



December 2000

# The Process of Physical Fitness Standards Development

		Job Performance		
		Can Do	Cannot Do	
Fitness Test	Pass	<b>A</b> True Positives	False Positives	<b>A +</b> Positive predictive value = $\frac{A}{A + \text{False Positives}}$
	Fail	<b>C</b> False Negatives	<b>D</b> True Negatives	<b>C + D</b> Negative predictive value = $\frac{D}{C + D}$
		<b>A + C</b>	<b>+ D</b>	
		Sensitivity = $\frac{A}{A + C}$	Specificity = $\frac{D}{+ D}$	

Distribution Statement A:  
Approved for public release;  
distribution is unlimited.



**Editors:**  
Stefan Constable  
Barbara Palmer

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 074-0188	
<small>Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.</small>				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE 18 December 2000		3. REPORT TYPE AND DATES COVERED State-of-the-Art Report 18 December 2000
4. TITLE AND SUBTITLE The Process of Physical Fitness Standards Development			5. FUNDING NUMBERS SPO700-98-D-4001	
6. AUTHOR(S) Constable, S.H. and Palmer, B. (Eds.)				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) HSIAC Program Office AFRL/HEC/HSIAC, Bldg. 196 2261 H Street WPAFB, OH 45433-7022			8. PERFORMING ORGANIZATION REPORT NUMBER HSIAC-SOAR-2000-001	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Performance Enhancement Division USAF School of Aerospace Medicine 311th Human Systems Wing Brooks Air Force Base, TX 78235-5252			10. SPONSORING/MONITORING AGENCY REPORT NUMBER Defense Technical Information Center ATTN: DoD IAC Program Office (DTIC-AI) 8725 John J. Kingman Road, Suite 0944 Fort Belvoir, VA 22060-6218	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.				12b. DISTRIBUTION CODE A
14. SUBJECT TERMS physical fitness; standards; cut-point; health based fitness; physical fitness standards; occupational specialties; physical demands				15. NUMBER OF PAGES 288
				16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED		20. LIMITATION OF ABSTRACT Unlimited

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2.89)  
Prescribed by ANSI Std. Z39-18  
298-102

# Contributors

**Stefan Constable, Ph.D.**

Chief, Performance Enhancement Division  
USAF School of Aerospace Medicine  
2602 West Gate Road  
Brooks Air Force Base, TX 78235-5252  
Stefan.Constable@brooks.af.mil

**Deborah L. Gebhardt, Ph.D.**

Human Performance Systems, Inc.  
5000 Sunnyside Avenue, Suite 203  
Beltsville, MD 20705  
hpsdgebhardt@erols.com

**James A. Hodgdon, Ph.D.**

Human Performance Center  
Naval Health Research Center  
P.O. Box 85122  
San Diego, CA 92186-5122  
hodgdon@nhrc.navy.mil

**Andrew S. Jackson, P.E.D.**

Department of Health and Human Performance  
University of Houston  
Houston, TX 77004  
Ajackson@jetson.uh.edu

**Barbara Palmer**

AFRL/HEC/HSIAC  
2261 Monahan Way, Building 196  
Wright-Patterson Air Force Base, OH 45433-7022  
Barbara.Palmer@wpafb.af.mil

**Mark Rayson, Ph.D.**

Optimal Performance Limited  
Old Chambers  
93/94 West Street, Farnham  
Surrey GU97EB, United Kingdom  
Mark@optimalperformance.co.uk



# **The Process of Physical Fitness Standards**

*Edited by*

Stefan Constable

USAF School of Aerospace Medicine

Brooks Air Force Base, TX 78235-5252

Barbara Palmer

Human Systems Information Analysis Center

Wright-Patterson Air Force Base, OH 45433-7022

Human Systems Information Analysis Center

Wright-Patterson Air Force Base, Ohio, 2000

## Description of cover design

The real goal of any testing standard approach is to maximize the correct predictions of (job) success or failure (cells A or D) and minimize the incorrect assessments (cells B and C). Various statistical tools are available to assist in the evaluation of test accuracy. This is a notional example of a contingency table and is analogous to the classical truth table used in testing the null hypothesis ( $H_0$ ) for Type I (a) and Type II (b) error. As the illustration indicates, factors such as sensitivity, specificity, and predictive values covary with the ratio of correct decisions to the total number of possible decisions. Overall, this notional depiction well represents a major theme of this SOAR—the primary goal of developing valid test standards can become surprisingly perplexing.

# Preface

---

Over the years, there has been considerable interest both in job performance and standards for physical fitness, especially within the Armed Services. Unlike most of the learned scholars who have contributed to this compendium, my association and interest in performance/fitness standards are more recent. However, as my exposure to this field progressed, I became more and more intrigued, both with the supporting science, as well as the practical implementation of the fitness standards process itself. This interest and quest for more information ultimately led to the evolution of this present State-of-the-Art Report (SOAR).

Most of us have a general acceptance regarding a level of required mental performance for entry or advancement in the formal education process. And as pointed out early in this book, various organizations have sought to establish both mental and physical standards for job candidate assignment and worker retention, in order to ensure job performance and safety. The processes, however, are hardly as straightforward or black-and-white as most might envision. In fact, it has been recommended that perhaps both good science and good judgment are required in equal measure. Of course, the incorporation of standards for acceptable performance on tests of physical capacity should be scientifically (and legally) defensible, which is sometimes a lofty goal! However, the "rigors of review" may sometimes be relaxed in the military environment. For example, performance on current military fitness tests often does not correlate well with task-specific job requirements, especially those involving muscular strength, i.e., heavy lifting or carrying. Yet relatively few steps have been implemented by the services to update and improve their approach to the fitness standards process, particularly with regard to occupational and health considerations. Therefore, an underlying theme in this text is that the scientific process to establish defensible standards can be complex, and subject to interpretation and challenge. More specifically, it may not always be possible to achieve the desired degree of test validity. For example, cost-benefit concerns and resource considerations may be overly constraining to the expected outcomes. So in practice, the process is often varied or abbreviated and may at times, involve rather arbitrary decisions.

This reference document strives to provide an organized base of knowledge concerning the primary issues of standards development as they might apply to both the military and civilian environments. Much of the material is technically in-depth, but certain segments lend themselves to larger target audiences as well. Chapters written by academic, industry, or military subject-matter experts cover all of the key topics which are relevant to this field. These specific, contributing authors bring a unique breadth of background and experience from the academic, military, industrial and laboratory fields. I truly believe this review is the only one of its kind published to date.

Stefan Constable

This publication is copyrighted. Most of the illustrations were provided by commercial publishers and those companies retain the right to the individual figures and associated material.

This material may be reproduced by or for the U.S. Government pursuant to the copyright license under the clause at 52.227-7013 (May 1987).

Additional copies of this state-of-the-art report (SOAR) are available for US \$45<sup>00</sup> (plus shipping and handling) from—

Human Systems IAC Program Office  
AFRL/HEC/HSIAC  
2261 Monahan Way, Building 196, Room 8  
Wright-Patterson AFB, OH 45433-7022

Tel: (937) 255-4842  
Fax: (937) 255-4823  
E-mail: [sales@wpafb.af.mil](mailto:sales@wpafb.af.mil)  
URL: <http://iac.dtic.mil/hsiac>



# About Human Systems Information Analysis Center

The Human Systems Information Analysis Center (HSIAC) is the gateway to worldwide sources of up-to-date human-factors information and technologies for designers, engineers, researchers, and human-factors specialists. HSIAC provides a variety of products and services to Government, industry, and academia promoting the use of ergonomics in the design of human-operated and manned systems.

HSIAC's primary objective is to acquire, analyze, and disseminate timely information about ergonomics. On a cost-recovery basis, HSIAC will perform the following functions—

- Distribute human-factors and ergonomics technologies and publications.
- Conduct customized bibliographic searches and reviews.
- Prepare state-of-the-art reports and critical reviews.
- Conduct specialized analysis evaluations.
- Organize and/or conduct workshops and conferences.

HSIAC is a Department of Defense Information Analysis Center sponsored by the Defense Technical Information Center, Fort Belvoir, Virginia. It is technically managed by the Air Force Research Laboratory Human Effectiveness Directorate, Wright-Patterson Air Force Base, Ohio, and operated by Booz-Allen & Hamilton, McLean, Virginia.





# Foreword

---

Requirements or standards for physical fitness within the Armed Services have received considerable attention and focus in recent years. Reasons for this include advancement in the physiology and medicine underlying physical fitness, expanded research on the development and practice of fitness, expanded numbers of women in the military, and the shift in the role that physical fitness plays in military occupations and missions. Despite this enhanced focus on military physical fitness, relatively few steps have been implemented by the services to update and expand fitness standards, particularly in regard to occupational and health needs. The reluctance of the training and personnel policy communities to expand fitness requirements for military occupations may in part be due to the lack of an organized and published base of knowledge concerning the primary issues of standards development as they would apply to the military services. This landmark report finally responds to this important need by presenting authoritative chapters on the key topics of physical standards development. Each chapter, written by foremost military and industry experts in their specialties, presents an in-depth analysis of each major aspect that challenges both military and industrial developers of fitness requirements. Each topic, from job analysis to legal issues, provides both a theoretical basis as well as a practical guide to developers and policy makers of physical fitness requirements, both military and industry. This report, by bringing together under one cover all critical topics regarding the setting of standards, will make a significant contribution to implementing the proper role of physical fitness to occupational classification and utilization and is recommended reading to all those working in the various disciplines of physical fitness development.

**James A. Vogel, Ph.D.**

Retired, Former Director of Occupational Health & Performance

U.S. Army Research Institute of Environmental Medicine



# Executive Summary

---

This State-of-the-Art Report (SOAR) documents the methods, processes, and issues that are involved with the development of physical fitness standards with special reference to the military. There has been a long-standing interest in standards for individual physical fitness. On close examination, this topic is truly complex. This SOAR strives to provide an organized base of knowledge concerning the primary factors of standards development as they might apply to both the military and civilian environments. Chapters, each written by a military or civilian subject-matter expert, focus on history of occupational demands assessment, health-based fitness standards, job analysis, types of physical fitness tests, test validity, setting performance standards, and legal issues. This review is unique in both its scope and timeliness of information.

Knowing how fit personnel should be is one focus of the field of occupational demands measurement—a field that has its roots in the fields of industrial engineering and occupational assessment, individual differences, and physical fitness. Military fitness, or the lack of it, has been an issue for as long as militia have existed. The physical fitness programs of the military services now employ norm-based standards—those sometimes representative of past military populations. In terms of occupational fitness standards, only the Air Force currently employs such an occupational fitness test, the Strength Aptitude Test, which is administered at accession. Although military researchers have vigorously studied this topic, few steps have been implemented by the services to update and improve their approach to the fitness standards process, particularly with regard to occupational and health considerations. Further policy and programmatic changes will be necessary as the military services move into occupationally based fitness standards. On the other hand, industry and other organizations have progressed relatively further here, albeit rigorous legal constraints and challenges.

A more generic approach to standards development is to establish health-based fitness levels. The scientific literature is replete with publications supporting the strong association between physical activity/fitness and general health, wellness and quality of life (detailed in Appendix A). Increased life span, enhanced quality of life, and reduced morbidity/mortality result from an active and fit lifestyle. From a military standpoint, “generalized” fitness standards promote physical readiness commensurate with an active life style, decreased **risk** of injury and the deployability of the military profession. Health-based standards are therefore forwarded here as an adjunctive or baseline approach to the typical (performance-based) development process. One must first attempt to quantify the relationships between the exercise regimen (or  $R_X$ ), the measured level of fitness, and health benefit outcome. While the rationale of this approach is most clear, in some cases pragmatic barriers are evident and, for example, include insufficient data to identify clearly defined cut-points on which to base specific standards. Basically this stems from the difficulty in identifying *minimal* dose-response (versus *adequate* or *optimal*) relationships across the varying exercise modalities. Nevertheless, this should not deter further developmental efforts to apply or investigate alternative procedures for health-based (or baseline) fitness standards approaches.

Establishing the more classical, job-specific fitness standards is a process that should begin with a job analysis or Physical Demands Analysis, to describe and quantify those aspects of physical fitness or physical performance that are relevant to job performance. A number of techniques are available to identify the most physically demanding tasks using some industrial/organizational psychology tools, and to quantify the stress and strain associated with these tasks using physiological, biomechanical and psychophysical approaches. Each technique has its own strengths and limitations, and it is the responsibility of the investigator along with a host of other considerations, including resource availability, to develop the approach which is most likely to elicit a complete and balanced output. Overall, conducting a job analysis is a confound process that requires considerable investment. The best science and good judgement are required in equal measure. The output of a conscientious Physical Demands Analysis should provide a sound foundation for establishing occupational fitness standards, focusing physical training programs, identifying health and safety issues and prioritizing those tasks that require job redesign. The long-term benefit to the employer of implementing these strategies will be increased productivity through improved operational effectiveness and reduced injury.

Employing appropriate types of tests to evaluate an individual's ability to do specific, physically demanding work is a second essential step in developing rational fitness standards. There are two general types, basic ability and work sample tests. The basic ability tests (field or laboratory) include those that assess aerobic fitness, body composition, muscle strength, and muscle endurance (flexibility, agility and balance are normally physical factors of less importance). A work-sample test is designed to duplicate the occupational task. Whereas, basic ability tests are of lower order or more generic and normally require less skill or coordination. Research documents that physically demanding work-sample test performance is largely dependent on aerobic fitness, body composition and strength to varying degrees. When basic ability tests are highly correlated with work sample tests, one test administration option is to replace the work sample test with a basic ability test or combination of basic ability tests. This is especially desirable when approaching testing for job selection purposes.

The sanctioned test validation methods are content validity, criterion-related validity, and construct validity. At the risk of over simplification, one might relate these methods with work sample tests, basic ability tests and a weighted combination of basic ability tests, respectively. Important to the issue of test validity are the EEOC guide-lines and the American Psychological Association standards for validating educational and psychological tests. A major difference in physical test validation is the use of physiological rather than psychological tests. The goal of physiological validation is to define the physical or physiological capacity needed by a worker to perform the work demanded by the task. Principal features of a physiological validation approach are the use of ergonomic metrics to quantify test performance, along with the interpretation of validity results with relevant physiological research and theory. Examples of validation studies are also reviewed for: outside crafts, firefighters, highway patrolmen, steel workers, coal mining, chemicals production, electrical lineworkers, underwater divers, military and manual lifting avocations.

In the employment setting, test scores may be used to determine and predict acceptable job performance. Crucial to standards establishment is a rational methodology to establish passing test scores that identify individuals who are able to perform, or be trained to perform, the essential job tasks. These methodologies depend on the data generated when content and criterion-related validity strategies are used to identify legally defensible passing scores. Effective criterion measures

assess and differentiate levels of job performance, and these data, along with test scores and validity coefficients, are used to formulate passing scores that identify successful and unsuccessful candidates. Methods to assess whether a passing score maximizes correct testing decisions, while minimizing testing errors include expectancy tables, contingency tables, and Taylor-Russell tables. The goal is to maximize correct testing decisions, while also minimizing testing error. Issues related to the computation of test fairness and utility, and adverse impact, and their integration with legal considerations are also described. For example, test fairness and adverse impact on specific populations of workers are very important considerations of physical testing standards, along with the legal implications in many employment contexts. The effect of basic physiological tests (e.g., aerobic capacity, strength tests) and job simulations on the reduction of adverse impact is also shown using comparisons from a variety of physically demanding jobs.

Legal forces and issues related to employment practices include Title VII of the Civil Rights Act of 1964, the Age, Discrimination in Employment Act (ADEA) of 1967, and Americans With Disabilities Act (ADA) of 1990. The centerpiece of employment discrimination law is Title VII of the Civil Rights Act of 1964, as amended by Congress on several occasions. Title VII prohibits employment discrimination because of “race, color, religion, sex, and national origin” by employers, labor organizations, and employment agencies. Title VII tends to be comprehensive in that everyone is potentially covered, because both genders and all majority and minority racial and ethnic groups, as well as religious groups, are covered by Title VII, but the act does not apply to military personnel. The disparate impact theory is used to establish employment discrimination. This legal process has a three-part burden of proof. First, the plaintiff (employee) must establish that the hiring practice has a disparate impact on a protected group. Although not legally mandated, the EEOC Guidelines are often used to define disparate impact. The Guidelines use the four-fifths (4/5s) rule, i.e., less than 80%, of the pass rate of the group with the highest pass rate, to define adverse impact. Once adverse impact is established, the burden of proof then falls on the defendant (employer) to justify that the exclusionary effect is a business necessity. The defendant must show that the selection method is job related. A common method used to establish job relatedness is with a validation study. Lastly, if business necessity is established, the burden of proof shifts back to the plaintiff to demonstrate that the employer failed to use a selection device that is equally effective but has a lesser disparate impact. Many of these cases reviewed involve the use of height and weight standards and tests for selecting public service employees such as police officers and firefighters. The outcome of this litigation largely depended on the scientific quality of validation study. The recent court ruling of *Lanning v. SEPTA* (U.S. 3<sup>rd</sup> Circuit 1999) will likely have a major impact on physical testing, and strongly suggests that validation studies will not only be evaluated by standard psychometric criteria, but also physiological validation of the test and cutscore.





# Acknowledgmen

---

I need to thank so many people and organizations for their participation in this effort—first Dr. Stefan Constable, whose identification and elucidation of these important issues were the impetus for this State-of-the-Art Report, and to the Air Force Materiel Command for providing funding to the Human Systems Information Analysis Center. To all our SOAR authors and reviewers, for your technical expertise, scheduling flexibility, willingness to help in the midst of already burgeoning workloads, the biggest of thank yous—you made this project a pleasure.

Thanks to Tom Metzler, HSIAC Director; Mike Fineberg, Ph.D., HSIAC Chief Scientist; and Joe McDaniel, Ph.D., the Contracting Officer's Technical Representative for HSIAC, for their leadership and technical insight.

Thanks also to Beth Rogers, Marcy Birch, Mike Rench, Michelle Dahle, Jeff Landis, Heather Williams, Lori Mansfield, Gilbert Bandry, and Marie Palmer for invaluable research assistance and editorial support, and to Ahnie Senft and Christina McNemar for their impressive graphics and production skills. This has been a great team project!

Barbara Palmer

This editor would like to briefly note some special contributions in addition to each chapter author, without which this SOAR would not have been possible. First and foremost I would like to recognize the particularly heroic efforts of my co-editor, Ms. Barbara Palmer. Clearly her belief and enthusiasm in this project from the beginning has been paramount to both its quality and completion. Throughout this journey her hallmark approach has been the difficult balance of professionalism, persistence and politeness. I would also be remiss if I did not note the special expertise and efforts of Dr. Tony Jackson, clearly one of the elder statesman in this field. In addition to initial benchmark work on just one chapter contribution, Tony admirably and willingly stepped up to further contributions, over a short time course, when unexpected technical shortfalls evidenced themselves. The special interest and effort of the government technical manager, Dr. Joe McDaniel, also needs mention. As an experienced investigator in this area, Joe spent extra time to also review this work and add many helpful suggestions. Finally, I need to offer special thanks to Dr. Jim Vogel, for his willingness to help with the review and pen a very gracious Foreword to this work. Jim spent a good deal of his civil service career investigating these issues for the Army. Jim's landmark contributions to the field were in no small part recognized by ultimate promotion to the government's elite, Senior Executive cadre. Clearly there are others whose contributions have been critical to this publication and please accept my deepest thanks and appreciation.

Stefan Constable



# Table of Contents

## Chapter 1

### **History of Occupational Demands Measurement and the Services' Physical Fitness Programs 1**

---

Abstract .....	1
Introduction .....	1
History of Occupational Demands Measurement .....	2
Occupational Assessment .....	2
Measurement of Individual Differences .....	3
Physical Fitness Measurement and Military Fitness .....	4
Physical and Occupational Fitness Programs of the U.S. Military Services ..	6
Physical Fitness Programs .....	6
Occupational Standards Programs .....	7
History of U.S. Army Fitness Program .....	9
Current U.S. Army Fitness Program .....	10
History of U.S. Air Force Fitness Program .....	11
Current U.S. Air Force Fitness Program .....	12
History of U.S. Navy Fitness Program .....	14
Current U.S. Navy Fitness Program .....	15
History of U.S. Marine Corps Fitness Program .....	16
Evaluation of Military Fitness Standards .....	17
Evaluation of Military Body Composition Standards .....	19
Evaluation of Occupational Fitness Standards .....	20
Summary .....	20
References .....	21

## Chapter 2

### **Health-Related Fitness Standards: A Baseline Approach 27**

---

Abstract .....	27
Health and Fitness Across the Workplace .....	28
General Health and Fitness Descriptors .....	28
Metabolic Fitness .....	29
The Relationships Between Physical Activity, Fitness, and Health .....	30
Specific Impact of Activity and Fitness on Personal Health .....	32

Interrelated Impact on Conditions .....	33
Reduced/Increased Risk of Injury .....	34
General Sense of Well-Being .....	36
Everyday Work Performance and Cognition .....	36
Medical Costs and Health Benefits .....	37
Mortality and Level of Fitness .....	37
Conclusion .....	38
Exercise Prescription .....	38
Programs and Recommendations Regarding Physical Activity .....	40
American College of Sports Medicine Position Stand .....	40
Recommendations from Healthy People 2010 .....	41
Recommendations from CDC/ACSM .....	42
Recommendations from the American Heart Association .....	43
Recommendations from the National Institutes of Health .....	43
Comparison of the Recommendations of National Fitness and Health Groups .....	44
The Dose-Response Relationship and Test Outcomes .....	45
Physical Activity Versus Test Outcomes/Standards .....	47
Specific Approaches to Setting Cut-points or Standards .....	47
Aerobic .....	47
Muscle Strength and Endurance .....	48
Body Composition .....	49
Further Military Relevance .....	49
Readiness and Fitness .....	49
Job-Specific Fitness Requirements .....	50
U.S. Air Force Direction .....	51
Appearance in the Military Context .....	51
Summary .....	54
Endnote .....	57
References .....	58

## Chapter 3

### Job Analysis

**67**

---

Abstract .....	67
Introduction .....	68
What Is a Job Analysis? .....	68
Who Should Conduct a Job Analysis? .....	69
Which Employees Should be Evaluated? .....	69
Identifying the Physically Demanding Tasks and Elements .....	70
General Information Gathering .....	70
Task, Environmental, and Human Factors .....	71
Observation .....	73

Questionnaire .....	75
Worked Example 1 .....	76
Interview .....	77
Approaches to Quantifying the Physical Demands of Jobs .....	77
Physiological Approach .....	78
Worked Example 2 .....	83
Worked Example 3 .....	85
Biomechanical Approach .....	87
Psychophysical Approach .....	91
Worked Example 4 .....	91
Summary .....	93
References .....	94

## Chapter 4

### **Types of Physical Performance Tests 101**

Abstract .....	101
Introduction .....	101
Physical Ability Tests .....	102
Aerobic Fitness .....	102
Body Composition .....	109
Muscular Strength .....	116
Muscular Endurance .....	119
Flexibility .....	120
Work Sample Tests .....	121
Relationship of Basic Abilities on Work Sample Tests .....	122
Role of VO <sub>2</sub> max .....	123
Role of Strength and Fat-Free Weight .....	123
Need for Further Research .....	125
Relationship of Basic Abilities on Health .....	128
Aerobic Fitness and Health .....	128
Body Composition and Health .....	128
Strength and Injury .....	130
Summary .....	130
Endnotes .....	131
References .....	131

## Chapter 5

### **Physical Test Validation for Job Selection 139**

Abstract .....	139
Employment Selection Tests .....	139

Validity of Selection Tests .....	140
Requirements for Validity Studies .....	144
Physiological Validation .....	146
Differences in Physiological and Psychological Test Validation .....	147
Physiological Validation—Test Fairness .....	148
Physiological Validation—Cut-Score .....	152
Physiological Validation—Matching the Worker to the Job .....	158
Published Validation Studies .....	160
Outside Craft Jobs .....	160
Firefighters .....	161
Highway Patrol Officers .....	163
Steel Workers .....	163
Underground Coal Mining .....	164
Chemical Plant Workers .....	165
Electrical Transmission Lineworkers .....	166
Diver Training .....	166
Demanding Military Jobs .....	167
Manual Lifting Tasks .....	168
Summary .....	171
Endnotes .....	171
References .....	172

## Chapter 6

### **Establishing a Minimum Standard 179**

Abstract .....	179
Introduction .....	180
Requirements for Identifying Passing Scores .....	180
Methods for Determining Passing Scores or Standards .....	181
Formulation of Criterion Measures .....	182
Identification of Acceptable Job Performance .....	184
Decision Models for Investigating Test Effectiveness and Setting Passing Scores .....	185
Normative Data .....	191
Ergonomic Data .....	192
Test Fairness and Adverse Impact .....	193
Adverse Impact of Physical Tests .....	193
Test Fairness .....	194
Summary .....	196
References .....	197

## Chapter 7

### Legal Issues

**201**

---

Abstract .....	201
Introduction .....	202
Legal and Legislative Forces .....	203
Legal Process — Discrimination Litigation .....	204
Disparate Impact .....	204
Business Necessity—Job Related .....	206
Alternative Selection .....	207
Court Rulings .....	208
Height and Weight .....	209
Physical Tests .....	209
Lifting .....	211
Physiological Parameters .....	211
A View to the Future .....	217
Endnotes .....	218
References .....	218

## Appendix A

### Physical Fitness and Specific Health Outcomes

**223**

---

Overweight and Obesity .....	223
Hypertension .....	226
Coronary Heart Disease and Cardiovascular Disease .....	228
Diabetes .....	230
Osteoporosis .....	231
Cancer .....	232
Clinical Depression .....	234
Stroke .....	235
Musculoskeletal Problems .....	237
References .....	238

## Index

**243**





# List of Figures

Figure 2.1	Relationship among habitual physical activity, health-related fitness, and health status. ....	31
Figure 2.2	The dose-response curve represents that best estimate of the relationship between physical activity (dose) and health benefit (response). The lower the physical activity status the greater will be the health benefit associated with a given increase in physical activity (arrows A, B, and C). ....	40
Figure 2.3	Change in maximal oxygen uptake at onset of blood lactic acid for subjects after 12 weeks of endurance training. ....	46
Figure 2.4	Notional depiction of health-based physical fitness standards (Tier I) ....	52
Figure 2.5	Notional depiction of performance-based physical fitness standards (Tier II) ...	53
Figure 3.1	Illustration of the oxygen deficit at the onset of activity and the excess post-activity oxygen consumption (EPOC), indicating the interaction of aerobic and anaerobic energy pathways required to meet energy demands. ....	79
Figure 3.2	Trace of deep body temperature throughout a day's patrolling activity in a soldier, measured via an ingested telemetry pill. ....	86
Figure 3.3	Copyright 1995, From Wilson and Corlett's body map by J. R. Wilson & E. N. Corlett. ....	93
Figure 4.1	Scale for rating level of physical activity. The directions are to select one value that best represents the level of physical activity for the previous month. The scale was developed for use in the Cardio-pulmonary Laboratory, NASA/Johnson Space Center, Houston, Texas. ....	108
Figure 4.2	Curves of the relationship between aerobic fitness and health show that the curves level off at 32 ml/kg/min for women and 35 ml/kg/min for men. These values define the threshold level of fitness needed for health promotion (i.e., reduced mortality) for 45-year-old men and women. ....	110
Figure 4.3	The participants in the lowest aerobic fitness group had the highest death rates. The mortality rate of the moderate and highly fit group was about the same. The high mortality rate occurs mainly in the low fitness group. ....	129
Figure 5.1	Maximum box weight lifted to elbow height as a function of arm curl 1RM ....	142
Figure 5.2	Three experimental designs for construct validity studies. Single-ended arrows indicate variables included in the factor analysis. Double-ended arrows indicate correlations to be measured. ....	144

Figure 5.3	Test for fairness, example of significant differences in male and female regression slopes .....	.150
Figure 5.4	Test for fairness, example of parallel regression slopes, but significant differences in male and female regression intercepts .....	.151
Figure 5.5	Test for fairness, example of homogeneity of male and female regression slopes and intercepts .....	.152
Figure 5.6	Reprinted, by permission, from Altman, D. G., Blandm J. M. (Altman-Bland plot of prediction residuals (measured – estimated) contrasted by the average of measured and estimated maximum push force) ...	.154
Figure 5.7	Probability of being able to generate 100 pounds for push force for levels of strength .....	.155
Figure 5.8	Scatterplot of strength test of subjects who could or could not complete a 90-pound lift from floor to knuckle height .....	.157
Figure 5.9	Logistic curves of the probability of being able to lift the weight load as a function of lifter strength .....	.158
Figure 6.1	Graphic rating scale examples .....	.182
Figure 6.3	BOS example .....	.183
Figure 6.2	BARS scale example .....	.183
Figure 6.4	Examples of contingency table .....	.189
Figure 6.5	Basic ability tests: Women’s performance as percent of men’s. From “Applying regression analysis and biomechanical modeling to setting of cut-scores or qualifying standards,” by D.L. Gebhardt, 1999, paper presented at American College of Sports .....	.194
Figure 6.6	Job simulation tests: Women’s performance as percent of men’s. From “Applying regression analysis and biomechanical modeling to setting of cut-scores or qualifying standards,” by D.L. Gebhardt, 1999, paper presented at American College of Sport .....	.195
Figure 6.7	Differential prediction analysis .....	.196

# List of Tables

Table 1.1	Comparison of basic characteristics of the physical programs of the four services . . .	7
Table 1.3	Percent body fat standards for the military services and the U.S. Coast Guard . . . .	8
Table 1.2	Comparison of physical fitness assessment standards, adjusted for age and gender, across the four services (minimums) . . . . .	8
Table 1.4	Measurement sites for circumferential taping for body composition assessment . . .	9
Table 2.1	Components of motor performance, physical fitness, and health-related physical fitness . . . . .	30
Table 2.2	Aerobics Center longitudinal study health-related fitness standards . . . . .	48
Table 3.1	ACSM's Classification of physical activity intensity, based on physical activity lasting up to 60 minutes . . . . .	80
Table 4.1	American College of Sports Medicine standards for $\text{VO}_2\text{max}$ ml/kg/min . . . . .	109
Table 4.2	Age-adjusted adult aerobic fitness standards for health promotion above $\text{VO}_2\text{max}$ ml/kg/min . . . . .	110
Table 4.3	Effects of component errors on underwater determined by Siri percent body fat . . . . .	113
Table 4.4	U.S. Military body composition equations . . . . .	115
Table 4.5	Percent body fat standard for women contrasted by age group . . . . .	117
Table 4.6	Percent body fat standards for men contrasted by age group . . . . .	117
Table 4.7	A comparison of strength testing methods . . . . .	118
Table 4.8	The Pearson product-moment correlations between static and dynamic leg strength . . . . .	119
Table 4.9	Factor analysis of strength data published by Dempsey et al., 1998 . . . . .	120
Table 4.10	Correlations between the <b>sum</b> of Isometric Strength and Simulated Work Sample Tests . . . . .	125
Table 4.11	Product-moment correlations between body composition variables and $\text{VO}_2\text{max}$ expressed in relative (ml/kg/min) and absolute (ml/min) terms . . . . .	126
Table 4.12	Multiple regression analysis comparing the effect of fat-free weight and strength in combination with 1.5-mile run and firefighter work sample test performance . . . . .	127
Table 5.1	EEOC Technical standards Guidelines for the criterion validation method . . . . .	145
Table 5.2	EEOC Technical standards Guidelines for the content validation method . . . . .	146
Table 5.3	EEOC Technical standards Guidelines for the construct validation method . . . . .	146
Table 5.4	Sample sizes, strength means and standard deviations, and analysis of strength differences of those who could and could not lift the weight . . . . .	156
Table 6.1	Example of Taylor-Russell table for base rate .60 . . . . .	187
Table 6.2	Combined physical performance test battery expectancy table . . . . .	188
Table 6.3	Comparison of test score to acceptable and unacceptable job performance . . . . .	190

Table 7.1	Male and female pass rates for floor-to-knuckle height lifts and the application of the 4/5s rule .....	206
Table 7.2	Average $\text{VO}_{2\text{max}}$ (ml/kg/min) for males and females of selected ages. percentages that would pass the cut-score, and 4/5s rate for adverse impact .....	215





# Chapter 1

## History of Occupational Demands Measurement and the Services' Physical Fitness Programs

Barbara Palmer  
Human Systems Information Analysis Center  
Wright-Patterson AFB, OH

---

### Abstract

Knowing *how* fit Military personnel should be is one focus of the field of occupational demands measurement—a field that has roots in industrial engineering and occupational assessment, individual differences, and physical fitness. This chapter shows how interest in individual differences and physical fitness spans decades of time and a variety of focuses, and how the measurement of individual differences in physical abilities has long been intertwined with the needs of the Military. The physical fitness programs of the Military Services are documented, and an evaluation of the fitness standards reveals that a norm-based process has been used to establish most of these requirements. In the context of occupational fitness standards, only the U.S. Air Force employs such an occupational fitness test, the Strength Aptitude Test, which is administered at accession.

---

### Introduction

Military fitness, or the lack of it, has been an issue for as long as militia have existed. A low point in the state of Military fitness may have occurred in the United States during the Spanish-American War, when several obese U.S. Army generals were unable to mount their horses at the Battle of San Juan Hill in 1898. At that battle, a cavalry charge was led by a nephew of General Robert E. Lee, General Fitzhugh Lee, who was forced to travel into battle on the back of a donkey (DiDonato, 2000).

This chapter reviews significant milestones in the history of individual differences and physical fitness. For example, progress in the measurement of individual differences in physical abilities was

prompted by the needs of the Military—especially the manpower requirements of World Wars I and II. More recently, the requirement for equal opportunity in the workplace has led to new interest in establishing occupational fitness standards.

Establishing meaningful fitness standards is one focus of the field of occupational demands measurement—a field that encompasses knowledge-bases of several disciplines. In a chapter titled *Physical Abilities in the Handbook of Industrial & Organizational Psychology*, Hogan (1991) states that, “We have no formal history from which current efforts in physical abilities research evolved. Like the field of industrial psychology, roots can be traced to aspects of industrial engineering, applied psychology, and individual differences measurement. The common theme among these otherwise diverse fields is that of measurement...” (Hogan, 1991a, p. 755).

Following a brief history of the contributions of the fields of occupational assessment, individual differences, and measurement of fitness and Military fitness, this chapter summarizes the fitness programs of the United States Military Services, and describes the history and status of job-specific fitness standards in the Military.

---

## History of Occupational Demands Measurement

This section describes the major achievements of three fields of endeavor that have coalesced into the area of occupational demands measurement—occupational assessment, the study of individual differences, and fitness measurement and Military fitness—during the late 1800s to the present.

### Occupational Assessment

A progression of quantitative approaches to workplace activities has led to today’s interest in establishing occupational fitness standards. This progression began when pioneers in industrial engineering made an early impact on the measurement of workplace activities. Foremost among these early researchers was Frederick W. Taylor (Taylor, 1923), who espoused the practice of scientific management. Still a major tool for industrial engineers, his methodology for the study of scientific management was time study, which Taylor introduced to the Midvale Steel Company in 1881. Taylor employed a series of investigations to determine what made up a day’s work for a “first class man.” He timed the elements that composed a task, then translated the work done into foot-pounds or fractions of horsepower. His method required the precise timing of each of the smallest elements of a task, a determination of the quickest and most efficient motions to complete the operation, and computation of the amount of rest needed to perform each task. After these standard times were determined, Taylor was able to form the basis for training, performance measurement, incentives, and compensation.

Following Taylor’s scientific, quantitative view of the workplace were Lillian and Frank Gilbreth, who developed a related concept, that of motion study, in an effort to determine the best method for bricklaying (Gilbreth, 1901). Gilbreth knew that the various types and numbers of movements employed by various workers were different, and were associated with different levels



of output. He defined motion study as “the science of eliminating wastefulness from using unnecessary, ill-directed, and inefficient motions.” The principles of time and motion study remain much the same since their development by Taylor and Gilbreth.

The fact that these principles have endured is substantiated and illustrated in the following quote by Hogan (1991a)—

*Today's practicing industrial engineers receive training in time and motion methods, and Taylor and Gilbreth's procedures form the basis of task analysis and personnel training for jobs that require skilled motor performance. In addition, substantial research and application has led to the successful adoption of industrial performance standards, i.e., cut-points. Using time and motion analyses, standard performance times are adopted for performance of tasks within a system. Although performance standards serve many purposes, including performance appraisal, the standards provide an answer to the original and fundamental question of what constitutes a day's work. (p. 758)*

The greatest contribution to occupational assessment methodology during recent decades was Edwin Fleishman's work (1964). His research in human abilities and performance taxonomies provided a basis for the evaluation of Military and educational physical fitness tests. Fleishman's taxonomies were composed of attributes characterized as relatively unchanging “enduring traits” (1964, p.12). He conducted the first systematic research designed to determine the number of attributes necessary to generate an adequate taxonomy of physical performance. His nine basic physical abilities form three classes important to the structure of occupational fitness standards: muscular strength, cardiovascular endurance, and factors affecting movement quality.

Factor analysis of physical fitness dimensions including strength, speed, flexibility, balance, and coordination scores, led to recommendations that Military fitness tests include stamina and cardiovascular endurance factors, not previously included. A reanalysis of this work by Hogan (1991b) forms a classic study of work requirements and occupational performance.

## Measurement of Individual Differences

Parallel efforts in the field of applied scientific psychology during the late 1880s to the early 1900s were focused on understanding and measuring differences among individuals, a field known as psychometrics. In the late 1880s, Francis Galton gathered data on both physical and psychological characteristics (Thorndike & Hagan, 1969), being interested first in techniques of precise measurement and secondarily in the development of statistical procedures with which to make the data meaningful. Further development of statistical tools was accomplished by Karl Pearson and Charles Spearman. Their development of statistical techniques made possible the accurate analysis and description of patterns of individual differences (Thorndike & Hagan, 1969). Heavily influenced by Galton, James McKeen Cattell furthered the research in individual differences and wrote extensively about physical anthropometry and mental tests (Hogan, 1991a). The focus of both Galton and Cattell's work was human development, with the hope that assessment of individual traits could be used in educational and vocational arenas. Although two research studies (Sharp, 1898; Wissler,

1901) criticized Cattell's work and effectively ended his investigations, a new need for assessments of physical fitness and psychological fitness arose as the United States entered World War I.

Between World Wars I and II, there was a renewed interest in using tests and measurements for scientific and instructional applications (Hogan, 1991a). E.L. Thorndike, a student of Cattell's before the turn of the century, influenced the development of standardized educational tests. Alfred Binet and Lewis Terman produced several versions of intelligence tests, while motor testing was the focus of David K. Brace and Frederick Rand Rogers. Drawing on the foundation created by these leaders, as well as that established by Clifford Lee Brownell, Harvey Lehman, and Paul Andrew Witty, the decade before World War II experienced an intensive increase in scientific exploration and documentation that underlies much of the current practice (Hogan, 1991a).

## **Physical Fitness Measurement and Military Fitness**

The state of physical fitness assessment in the 1890s had consisted largely of three areas of measurement — anthropometry; muscular strength, endurance, and power; and cardiovascular fitness (Maud, 1995). Early sets of anthropometric measurements (Hitchcock & Seelye, 1893) included age, height, and weight; and chest, arm, and forearm girth. Seaver (1890) is considered the first modern author of fitness anthropometry. The second area of fitness measurement of this era was headed by Dudley Sargent, who is considered the father of modern strength testing, (Hartwell, 1885). Sargent's 1921 article, "The Physical Test of Man," described the vertical jump test, one of the first tests of muscular power. Kellogg (1896), who described the universal dynamometer that made this type of testing more accurate, conducted further work in the area of strength testing. Cardiovascular fitness testing in the 1900s was conducted with the Foster Cardio-Vascular Test, which correlated the response rate of the heart to exercise and recovery (Foster, 1914). The Barach Cardio-Vascular Test evaluated blood pressure and heart rate to determine cardiovascular function (Barach, 1919). Eventually, Schneider's (1920) more standardized combination of pulse rate and blood pressure assessment was used, taken with the subject both horizontal and standing, before and after exercise.

Although the methodologies for measurement were becoming more precise, the levels of fitness in the Military were not encouraging during this era, as illustrated at the beginning of this chapter. This low point in Military fitness levels spawned the need for assessments of physical and psychological fitness in the Military as the first World War loomed. The Commission on Training Camp Activities was formed to develop the physical capacities of World War I recruits, one third of whom were found physically unfit for duty, according to data cited by Bucher (1968). The situation was not much better on the civilian front, with a 1918 study of industrial employees showing that 270 million workdays were lost due to "loss of health and vigor" (Hackensmith, 1966, p. 412). These conditions led to legislation encouraging health and physical education in schools.

As the measurement of physical abilities began to improve, data were used to develop population norms. Rogers (1927) developed the Strength Index and the Physical Fitness Index, which consisted of seven measures, including hand, shoulder girdle, back, leg and arm strength, and forced vital capacity. The index was derived by comparison to norms based on gender, age, and weight; and its use led to great popularity of strength testing. Through the years, modifications to this test have

included the elimination of pull-up and dip tests, which were replaced with static tests of arm strength using the back/leg dynamometer (MacCurdy, 1933). Dynamometer testing of isometric strength is still used, but the more popular method is to assess maximum weight moved in one repetition (1RM). These indexes and tests measured the same three areas of physical abilities that were measured in the 1890s, but were updated and refined for the time span of the 1920s.

Cardiovascular fitness assessment of the early 1920s was characterized by Schneider's (1920) evaluation of World War I aviators. He combined pulse rate and blood pressure from subjects in the standing and horizontal positions, and assessed pulse rate after 15 seconds of bench stepping and during recovery. This method was followed by Tuttle's Pulse Ratio Test (1931), which assessed pulse rate before and after a prescribed period of bench-stepping activity. Currently, the Harvard Step Test (Brouha, 1943) is still used, or alternatively and perhaps more frequently, heart rates taken during and after exercise on a treadmill or a bicycle ergometer are used. The criterion measure for  $\text{VO}_2\text{max}$  is a maximal test that measures oxygen uptake, which can be done using sophisticated instrumentation and normally accomplished on a cycle ergometer or a treadmill.

Results of fitness tests made the lack of readiness of recruits entering World War I obvious. In response, the U.S. Army and U.S. Air Force set up a physical training course under baseball Hall-of-Famer Hank Greenburg; and the U.S. Navy set up a program under heavyweight boxing champion Gene Tunney. In 1942, the Office of Defense, Health, and Welfare Services developed a Division of Physical Fitness; and the Federal Security Agency established a committee on physical fitness in 1943. Because World War II statistics revealed that 900,000 out of 2 million volunteers were rejected because of "mental or physical defects" (Hogan, 1991), the necessity of fitness for all age groups became prevalent in the United States. The Minimum Muscular Fitness Test in School Children (Kraus & Hirshland, 1954) indicated that U.S. youth were less fit than their European counterparts. These results led to President Eisenhower's Presidential Conference on Fitness of American Youth in 1955, followed by the establishment of the President's Council on Youth Fitness, which continues to this day.

Continued concern for Military readiness occurred during the Korean Conflict and the Cold War. The concern about Military readiness was balanced with concern about the fitness of the American public, especially the youth. School children's fitness was assessed via the American Association of Health, Physical Education, and Recreation National Fitness Tests (AAHPER), which President Kennedy expanded into a fitness development program (Hunsacker, 1958).

In the 1960s on, there was a considerable push to increase the fitness levels of the American public. President Jimmy Carter commissioned a study to document the state of Military fitness in 1983 which led to body fat assessment programs being established by the Services (Institute of Medicine, 1998). During this decade and the next, increases in physical fitness testing and standardization were seen on many fronts. The American College of Sports Medicine led the way in making fitness a quantifiable, scientific study and published *Guidelines for Exercise Testing and Prescription* in 1991 and 1995. The YMCA Test Battery, the Canadian Standardized Test of Fitness (CSTF), the AAHPERD Test Battery, FITNESSGRAM, and the Eurofit Test Battery for Adults are all widely used tests. Norm-referenced standards were documented for several populations, including the Military. The Military standards that were derived during this era are described in the following section, and research efforts to increase ties to scientific findings over time are discussed.

---

# Physical and Occupational Fitness Programs of the U.S Military Services

## Physical Fitness Programs

Fitness standards in the Military have been established over the years to promote health and general physical fitness. General (or health-related physical fitness) standards are applicable to **all** Service members regardless of their occupation. These standards are not intended to specifically enhance the performance of a particular Service mission or job (see Chapter 2).

DoD's guidance, through Department of Defense Directive 1308.1, requires that the Services establish a physical fitness and body fat program that includes fitness requirements for all Service members (U.S. Department of Defense, 1995). This guidance requires that regardless of age, **all** Service members be measured along three dimensions annually: cardiovascular endurance (measured by activities such as running a certain distance within a specified time limit); muscular strength and endurance (measured by activities such as sit-ups and push-ups); and maintenance of body fat within a certain percentage range. The guidance does not specify particular testing activities or minimum required levels of difficulty. Each Military Service is required to design its own fitness program to meet mission-specific needs (U.S. Department of Defense, 1995).

Department of Defense Instruction 1308.3, Physical Fitness and Body Fat Programs Procedures (U.S. Department of Defense, 1995) applies to the Office of the Secretary of Defense, the Military Departments, the Chairman of the Joint Chiefs of Staff, and the Unified Combatant Commands. This policy states—

*It is DoD policy that physical fitness is essential to combat readiness and is an important part of the general health and well-being for Armed Forces personnel. Individual Service members must possess the cardio-respiratory endurance, muscular strength and endurance, and whole-body flexibility to successfully perform in accordance with their service-specific mission, and Military specialty. Those qualities, as well as balance, agility, and explosive power, together with levels of body composition, form the basis of the DoD Physical Fitness and Body Fat Program (p. 1).*

This same Instruction defines fitness as—

*The ability of Service members to meet the physical demands of their jobs for an extended period of time and to have the additional ability of meeting physical emergencies, such as those imposed during combat or other stressful situations.*

To help illustrate the qualities and the physical fitness guidelines outlined in the directive above, Table 1.1 compares the current fitness requirements of the U.S. Army, U.S. Air Force, U.S. Navy, and U.S. Marine Corps. Table 1.2 outlines the standards of the fitness tests used by the Services. Following this, Table 1.3 compares the body composition standards of the four Services, which

serve as a weight-for-height screen, and Table 1.4 indicates the components taken into account when the Services use circumference measurements to estimate body composition.

**Table 1.1 Comparison of basic characteristics of the physical programs of the four services**

	Army	Air Force	Navy	Marine Corps
<b>Reference</b>	Regulation 350-41, 600-9, and 600-63, FM 21-20	Instruction 40-501 and 40-502	instruction 611.1F	Order 6100.1c
<b>Objective/Goal</b>	Combat and operational readiness Healthy life style Military appearance	Motivation to train Fit and healthy force	Optimal health Stamina for optimal readiness	Overall fitness Mission/combat readiness
<b>Components</b>	Aerobic capacity Upper body/trunk strength/endurance Body fat	Aerobic capacity Upper body/trunk strength/endurance Body fat	Aerobic capacity Upper body/trunk strength/endurance Flexibility Body fat	Aerobic capacity Upper body/trunk strength/endurance Body fat
<b>Test Items</b>	2-mile run Push-up sit-up Body fat by tape	Submax cycle ergometer prediction of VO <sub>2</sub> max Push-up Ab crunch Body fat by tape	1.5-mile run/walk, or 500 yard swim Curl-up Push-up and reach Body fat by tape	3-mile run Ab crunch Pull-up (male) Flexed arm hang (female) Body fat by tape

Proposed addition. trial period through December 01.

## Occupational Standards Programs

The purpose of job-specific physical performance standards is to ensure that personnel assigned to physically demanding jobs can perform those jobs. Occupational fitness standards indicate the level at which an individual must perform in order to successfully meet job requirements, regardless of body size or gender. Military scientists have proposed approaches to quantify various task categories to determine the feasibility of establishing groups of fitness requirements that will be specific to job category.

Job-specific performance in the Military has been a concern since the U.S. Army Air Corps Aviation Psychology Program began in World War II (Hogan, 1991). During the last three decades, interest in job-specific performance testing has increased with the dramatic influx of women into the Military Services as well as into demanding occupational specialties. With increasing numbers of women entering the Military Services, in 1976, the GAO encouraged the DoD to develop physical standards for job performance using the Department of Labor system of classification (GAO, 1976, as cited in IOM, 1998). The GAO Guideline (Government Accounting Office, 1998a) report indicates that Section 543 of the 1994 National Defense Authorization Act, "Required the Secretary of Defense to prescribe physical performance standards for any occupation in which the Secretary determined that strength, endurance, and cardiovascular capacity was essential to the performance of duties" (p. 8). This act requires that, if developed, these standards will pertain to job activities that were commonly performed in that occupation, and must be relevant to successful physical performance of those tasks, and could not be based on gender. "In other words, job-specific physical performance standards will identify the absolute minimum level needed for successful performance in those occupations. Anyone in that occupation, regardless of gender, will be required to meet the same

**Table 1.2 Comparison of physical fitness assessment standards, adjusted for age and gender, across the four services (minimums)**

	Army			Air Force			Navy			Marine Corps		
	Age	Male	Female	Age	Male	Female	Age	Male	Female	Age	Male	Female
Aerobic Capacity	2-mile Run (min:sec)			Submaximal Cycle Ergometry (ml/kg·min $\dot{V}O_{2\max}$ )			1.5-mile Run/Walk (min:sec)			3-mile Run (min:sec)		
	17-21:	15:54	18:54				17-19:	12:30	15:00	17-26:	28 min	31 min
	22-26:	16:36	19:36	<24:	27	35						
	27-31:	17:00	20:30	25-29:	27	34	20-29:	13:30	15:30	27-39:	29 min	32 min
	32-36:	17:42	21:42	30-34:	27	32						
	37-41:	18:18	22:42	35-39:	26	31	30-39:	14:30	16:45	40-45:	30 min	33 min
	42-46:	18:42	23:42	40-44:	26	30						
	47-51:	19:30	24:00	45-49:	25	29	40-49:	15:30	17:15	46+:	33 min	36 min
	52-56:	19:48	24:24	50-54:	24	28						
	57-61:	19:54	24:48	55-59:	22	27	50t:	16:45	17:30			
	62+:	20:00	25:00									
Upper Body Strength/Endurance	Push-ups in 2 minutes						Push-ups in 2 minutes			Pull-ups (Males) Flexed Arm Hang (Females)		
	17-21:	42	19	Proposed addition of push-ups, in trial period through December 01			17-19:	42	19	17-26:	3	15 sec
	22-26:	40	17				20-29:	37	16	27-39:	3	15 sec
	27-31:	39	17									
	32-36:	36	15				30-39:	31	11	40-45:	3	15 sec
	37-41:	34	13									
	42-46:	30	12				40-49:	24	7	46+:	3	15 sec
	47-51:	25	10									
	52-56:	20	9				50t:	19	2			
	57-61:	18	8									
	62+:	16	7									
Abdominal Strength/Endurance	Sit-ups in 2 minutes						Curl-ups in 2 minutes			Ab crunches in 2 minutes		
	17-21:	53	53	Proposed addition of crunches, in trial period through December 01			17-19:	50	50			
	22-26:	50	50				20-29:	46	46	27-39:	45	45
	27-31:	45	45									
	32-36:	42	42				30-39:	40	40	40-45:	45	45
	37-41:	38	38									
	42-46:	32	32				40-49:	35	35	46+:	40	40
	47-51:	30	30									
	52-56:	28	28				50t:	29	29			
	57-61:	27	27									
	62+:	26	26									

**Table 1.3 Percent body fat standards for the military services and the U.S. Coast Guard**

Military Branch	Men		Women	
	Age (Years)	% Body Fat	Age (Years)	% Body Fat
Air Force	<29	20	<29	25
	30+	24	30+	32
Army	17-20	20	17-20	30
	21-27	22	21-27	32
	28-39	24	28-39	34
	40+	26	40+	36
Navy	17-39	22	17-39	33
	40+	23	40+	34
Marine Corps	All	18	All	26
Coast Guard	<30	23	<30	33
	31-39	25	31-39	35
	40+	27	40+	37

**Table 1.4 Measurement sites for circumferential taping for body composition assessment**

Military Branch	Men	Women
U.S. Air Force	Abdomen Neck	Waist Hip Neck
U.S. Army	Abdomen Neck	Wrist Neck Forearm Hip
U.S. Navy	Abdomen Neck	Waist Hip Neck
U.S. Marine Corps	Abdomen Neck	Waist Hip Neck

standard” (Government Accounting Office, 1998a, p. 8). Each Service has categorized its occupational specialties according to upper body strength, and the P-U-L-H-E-S physical profile (physical capacity or stamina, use of upper and lower extremities, hearing acuity, normal color vision (e), and special psychiatric characteristics) (Institute of Medicine, 1998).

In 1993, legislation that opened all MOSs (except direct combat) to women led to the GAO being directed to reopen the issues of job-specific performance testing. The GAO recommended that the Services determine whether a significant problem existed in the accomplishment of physically demanding occupations, and to identify ways to solve the problem. The GAO report (Government Accounting Office, 1996) recommended establishing valid performance standards, providing job training, and redesigning job tasks. Only the U.S. Air Force requires recruits to take a strength test for job assignment (Institute Of Medicine, 1998).

Presently, the U.S. Air Force AFSCs are categorized into eight physical demand categories, and the U.S. Air Force uses a strength aptitude test to screen out those recruits who would not be likely to perform successfully on a given job (U.S. Department of the Air Force, 1994). The U.S. Air Force does not incorporate this fitness test into the required annual fitness evaluation. The U.S. Navy has not adopted occupational strength standards for active duty personnel or recruits, nor has the U.S. Army. The U.S. Marine Corps at one time administered a physical readiness test of combat skills, but this has been discontinued.

## History of U.S. Army Fitness Program

**Physical Fitness** — Beginning in the middle of the nineteenth century, a European style of physical training was used in the U.S. Army. This fitness program was initiated by Herman J. Koehler, who was the “architect of modern U.S. Army physical readiness training” (<http://192.153.150.25/usapfs/pages/history1885%2D1920.htm>). He was a master of the German Turner System and until WW I, physical training was under his leadership. Near the end of this era, a new method of U.S. Army fitness was implemented, following the publication of the 1914 Manual of Physical Training for the United States Army. Between this time and the second World War, competitive athletics

was used as a supplement to conditioning for soldiers. The emphasis on motor skill development was aided through climbing, leaping, and tumbling. Many exercises were used for developing coordination and these included marching, running, dumbbell and club exercises, and jumping. Still other skills such as self-reliance and confidence were emphasized and stressed through swimming, climbing, boxing, wrestling, and gymnastic exercises (<http://192.153.150.25/usapfs/pages/history1885%2D1920.htm>).

During the U.S. involvement in WW II, the U.S. Army called on civilian and Military specialists to assist in the formation of a modern physical fitness program. This would be the first program of physical training that could be justified by accepted scientific testing procedures. Physical fitness was encouraged during WW II training camp with such expressions as, “the more sweat, the less blood” (<http://192.153.150.25/usapfs/pages/history1885%2D1920.htm>). The importance of strength and good health was stressed as well as physical conditioning that helped the soldier endure severe physical demands. Training to facilitate realistic combat situations was conducted with a soldier’s full pack, weapons, and ammunition. The formal calisthenics practiced during WWI were now being supplemented with a variety of difficult physical challenges.

After WWII, the U.S. Army consolidated physical training and athletics and revised the U.S. Army physical fitness publication FM 21–20, which was again revised in 1985, 1992, and 1998. During the 1960s, the need for a flexible response to a variety of threats led to an emphasis on combat physical training. President Carter’s 1980 call for renewed physical fitness efforts in the Military resulted in a new U.S. Army fitness initiative and the creation of the U.S. Army Physical Fitness School, established at Fort Benjamin Harrison and then moved to Fort Benning in 1997. The program now emphasizes physical training for the battlefield, health, diet, and other aspects of wellness (<http://www-benning.army.mil/usapfs/Doctrine/History/historykorea-present.htm>).

## Current U.S. Army Fitness Program

Physical Fitness—The U.S. Army conducts the U.S. Army Physical Fitness Test twice a year and a weight assessment once a year. Field Manual 21–20 directs the U.S. Army fitness program (U.S. Department of the U.S. Army, 1998). The U.S. Army Physical Fitness Test (Vogel, 1986) is administered to all personnel through age 60—individuals 40 years of age and older receive a pre-test physical and coronary risk assessment. The current set of tests were chosen on the basis of ease of administration, relative objectivity of scoring, and lack of needed equipment (Vogel, 1986). The three events include a timed 2-mile run, the maximal number of extended-leg push-ups within 2 minutes, and the maximal number of bent-knee sit-ups within 2 minutes. Acceptable substitutions for the 2-mile run are a 6.2-mile bike ride, a 2.5-mile walk, or an 800-yard swim. Scoring differs by age and sex.

In October 1998, the U.S. Army began to implement new standards that were again norm-based. This was done with the intent of maintaining a gender-neutral, equal-points-for-equal-effort standard. The new minimum scores are based at or near the 8<sup>th</sup> percentile of a sample of scores collected during a 1995 study, with maximums set at the 90<sup>th</sup> percentile. Standard scores are reduced in 5-year increments as age increases. The new standards increase some of the requirements for both sexes, requiring women to perform the same number of sit-ups as men. Female run



times are about 14 to 16 percent slower than male times, and female push-up requirements will increase from 44 to about 50 percent of the male standards (U.S. Department of the Army, 1998).

As with the other Services, the U.S. Army also has body composition requirements. Currently, the U.S. Army allows body fat of up to 20 to 26 percent for men and between 30 to 36 percent for women, depending on age. The lower figure for males was derived from data on young male soldiers from a decade ago (GAO, 1998a), adding 2 percent body fat for each decade after the second yields the upper figure of 26 percent. Up until 1991, female standards of 28 to 34 percent were derived by adding 8 percentage points to the male standards. These figures were determined to be more restrictive than the men's when the allowable body fat was compared to the means of same-sex recruits. In 1991, 30 and 36 percent were established as the lower and upper figures for U.S. Army women. The equations incorporating circumference measures used by the U.S. Army were developed by Vogel and coworkers (Vogel, Kirkpatrick, Fitzgerald, Hodgdon, & Harman, 1988) and include measurement of wrist, neck, forearm, and hip for females, and neck and abdomen for males.

For U.S. Army enlisted personnel and non-commissioned officers, physical training each morning is usually mandatory. For the most part, officers are responsible for their own physical conditioning program, although whether this is voluntary or mandatory varies from post to post.

**Occupational Fitness**—As many Military Occupational Specialties (MOSs) opened to U.S. Army women in the 1970s, the U.S. Army Research Institute for Environmental Medicine (USAFUEM) developed a battery of performance tests corresponding to capacities needed for the various MOSs, but unfortunately the tests were not applied. In 1981, USARIEM was again tasked with developing and validating a gender-neutral strength test to be administered as an entrance criterion. The resulting MEPSCAT (Military Entrance Physical Strength Capacity Test) was an incremental dynamic lift test, similar to the SAT, which predicted performance on job-related criterion performance tasks (Institute of Medicine, 1998). The study that validated the incremental lift test for this application was criticized for misanalyzing the data and overstating the correlation between lift performance and criterion test performance. The MEPSCAT was administered until 1990 and was used to counsel recruits about requirements of various physical occupations. Currently there is no U.S. Army occupational fitness test in place (Institute of Medicine, 1998).

## History of U.S. Air Force Fitness Program

**Physical Fitness**—The first U.S. Air Force publication on physical fitness was published in 1947 (U.S. Department of the Air Force, 1947). Although no standard program was documented, and no specific level of fitness was required, AFR 50-5 stated that the U.S. Air Force's fitness program should—

- \* *Develop and maintain a high level of physical fitness in the individual so that he can perform more efficiently his assigned duties*
- *Encourage regular and healthful exercise*
- *Foster an aggressive and cooperative team spirit, increase the confidence of the individual, develop sportsmanship, and increase pride through participation in competitive athletics. (cited in Schellhous, 1982, p. 14)*

Further guidance was included in U.S. Air Force Manual 160-26, *Physical Conditioning*, published in 1961 (U.S. Department of the Air Force, 1961). The manual stated that it was the commander's responsibility to see that his men were developed physically, psychologically, and socially. A standard program was still not established in this improved and updated document. A 1959 study (Balke & Ware, cited in Schellhous, 1982) of fitness levels in the U.S. Air Force found that the overall state was "poor," and declared that the U.S. Air Force physical fitness program was ineffective. AFR 50-5 was revised in 1959. Commanders were required to establish a physical conditioning program, establish weight limits, and prescribe regular weekly exercise. Still, no standard program or levels of fitness were established (Schellhous, 1982). The increased national interest in physical fitness in the early 1960s led the U.S. Air Force to adopt the Royal Canadian Air Force Five Basic Exercise (5BX) Plan as its official fitness program (U.S. Department of the Air Force, 1962). The U.S. Air Force Pamphlet (AFP) 50-5-1 regulated the program for men, and AFP 50-5-2 outlined the Ten Basic Exercise Plan (XBX) for women. The 5BX program was designed to progressively work the skeletal muscles, heart, and lungs until a given level of fitness was achieved. A specified number of repetitions of each of five exercises were to be completed within 11 minutes. In 1963, U.S. Air Force personnel and researchers from Indiana University reported that the 5BX program was fraught with problems, including a high failure rate, an unsatisfactory testing program, and a lack of emphasis on the importance of physical fitness (Schellhous, 1982).

Dr. Kenneth Cooper, a U.S. Air Force flight surgeon who is considered to be the founder of the aerobics movement, was behind the next iteration of the U.S. Air Force fitness program, in which an aerobics test was administered semi-annually to all personnel. The test consisted of a timed 1.5-mile run. Taking into consideration the person's age and the time taken to finish the run, personnel were put into one of five fitness categories ranging from very poor to excellent (U.S. Department of the Air Force, 1977). It was suggested that concern about fatalities during testing led the U.S. Air Force Surgeon General to modify AFR 35-11 in 1979. Personnel age 35 and over were tested using a 3-mile walk rather than the 1.5-mile run. This change was unpopular, and by 1980, all personnel were permitted to run rather than walk for the annual fitness test. In 1981, AFR 35-11 (U.S. Department of the U.S. Air Force, 1981) indicated that members to be tested annually could choose the 1.5-mile run, the 3-mile walk, or stationary running. In March 1982, the U.S. Air Force began assessing the cycle ergometry test to estimate work capacity, which led to the current U.S. Air Force cycle ergometry program which was evaluated during a ten year period and implemented in 1992.

## **Current U.S. Air Force Fitness Program**

**Physical Fitness** — The U.S. Air Force Medical Operations Agency established the U.S. Air Force Fitness Program Office in June 1995. The U.S. Air Force Fitness Program Office implements, sustains, and supports the U.S. Air Force Fitness Program (AFFP) for all U.S. Air Force entities. The U.S. Air Force Surgeon General's office is responsible for AFFP policy and procedures.

U.S. Air Force Instruction 40-501 (Medical Command, Department of the Air Force, 1996) outlines the U.S. Air Force Fitness Program, and states that cardiovascular (aerobic) fitness is the single best indicator of total physical fitness. The current cycle ergometry test was implemented in October 1992, and all U.S. Air Force personnel are tested annually. Cycle ergometry is used because

it is a reliable and safe estimate of cardiovascular fitness and is characterized by acceptable physiological validity. It is a submaximal test, based on the physiological principle that the heart rate increases as work intensity and oxygen consumption increase. The individual's heart rate and workload are used to estimate  $\dot{V}O_{2\max}$ . As an adaptation to physical training, heart rate is expected to decrease for a given level of work. This test has demonstrated its correlation with the graded treadmill test (Pollock et. al., 1994). The testing procedure basically assesses heart rate at the end of a 6 to 8 minute steady-state cycling period. Minimum, steady-state passing heart-rate-driven scores are established by sex and age. In its current iteration, the test is graded pass or fail only, with respect to population norms. An informal account of the U.S. Air Force's fitness history reveals that the cardiovascular standard was generally based on performance statistics from a population of U.S. Air Force men and women in the early 1990s (Government Accounting Office, 1998a).

Following a review of the literature on the benefits of strength training (Palmer & Soest, 1997), the U.S. Air Force is implementing a program that incorporates measures of strength/endurance (the ability to sustain submaximal contraction) and encourages flexibility activities. The expanded fitness program is being field tested in the years 2000 and 2001, with pass/fail standards to be implemented by January 2002.

The current U.S. Air Force fitness standard does not mandate duty time for physical fitness. Individual commanders are encouraged to require on-duty physical training, but for the most part, physical training is the responsibility of the individual.

In addition to aerobic fitness, the U.S. Air Force also maintains minimum standards of body composition. A history of the U.S. Air Force body composition program was derived by authors of the 1998a GAO report through discussions with officers who were responsible for the program. U.S. Air Force officials were unable to provide specific studies or records to document their body fat standards. The U.S. Air Force uses the same body fat estimation equations as the U.S. Navy's and measurements include circumference of waist and neck for men and abdomen, neck, and hip for women.

**Occupational Fitness**—Occupational fitness within the U.S. Air Force is part of the U.S. Air Force Personnel Center's classification program, and is not part of the U.S. Air Force Fitness Program, has focused on the Strength Aptitude Test (SAT). The SAT is a weight-lifting test performed at the Military Entrance Processing Stations (MEPSs) by all U.S. Air Force recruits under U.S. Air Force Policy AFMAN 36-2108 Attachment 39. The SAT was developed in five phases between 1977 and 1982 with the goal of matching abilities of recruits with the demands of individual jobs. The SAT was changed from a test program to a fully implemented program in 1987 (McDaniel, personal communication, 2000). The SAT uses an incremental lift machine that consists of a vertically moving carriage with handgrips. The weight to be lifted can vary from 40 to 110 pounds, in increments of 10 pounds. The handgrips are 16 inches apart and are 12 inches from the floor. The carriage and weight is lifted overhead. Lifts are repeated with increased weight until the lift cannot be safely completed at the 6-foot level or above the head.

During development of the SAT, several strength tests were considered as candidates. An incremental lift test to a height of six feet was finally selected because it proved to be the best single test for U.S. Air Force-wide use. After the test was developed and implemented, the empirical relationship between test performance and performance on specific job tasks was documented. The next phase was to survey the physical demands of every career field in the U.S. Air Force. This ambitious

project measured the physically demanding components, such as lifting and pushing of each job. The U.S. Air Force adopted strength requirements for all enlisted U.S. Air Force Specialties (AFSs) in 1988. The make-up and the duties of an AFS do not remain constant. Since the original survey of physical demands during 1978 to 1982, some AFSs have been subdivided and some combined to create new ones. In some cases, the systems and equipment used by the members of an AFS have changed. When changes in the job result in changes in the physical demands of the job, the SAT criterion may become obsolete. If this occurs, the physical demands must be reanalyzed to determine if the SAT criteria should be changed. Each year, it is the goal of the U.S. Air Force Research Laboratory's (AFRL) Human Engineering Directorate to resurvey several AFSs and compute new SAT criteria for them.

## History of U.S. Navy Fitness Program

**Physical Fitness** — The history of the U.S. Navy's fitness program is outlined in a technical report by Hodgdon (1999), who states that the first Naval Operations Instruction 6110.1 (Chief of Naval Operations, 1976) emphasized cardiovascular fitness, based on the popular aerobics program of Dr. Kenneth Cooper. The next version of this instruction, Naval Operations Instruction 6110.1A (Chief of Naval Operations, 1980), was the first to include a physical fitness test to allow assessment of each individual's level of fitness. Issued in 1980, it followed President Jimmy Carter's request for such an assessment among all the Military Services.

Naval Operations Instruction 6110.1B (Chief of Naval Operations, 1982) was issued to implement policies as directed by the Department of Defense in its Directive 1308.1 (U.S. Department of Defense, 1981). This instruction included physical fitness, weight control, and health promotion considerations, and outlined the Physical Readiness Test (PRT). The PRT consisted of a 1.5-mile run, or the number of steps-in-place that could be done in 3 minutes, the number of curl-ups that could be done in 2 minutes and measurement of the sit-and-reach flexibility range. In addition, it appointed Command Fitness Coordinators to manage the Commanding Officers for program implementation. Consequences for failing the PRT were documented, and two new items were added to the PRT—the number of push-ups that could be performed in 2 minutes and the time required to swim 500 yards or 450 meters (an alternative to the 1½-mile run). Ability to run in place was dropped as a test item, and the sit-reach was made a pass-fail item.

The next instruction, Naval Operations Instruction 6110.1C (Chief of Naval Operations, 1986), was limited to physical fitness and body fat standards, with health promotion being covered under a separate Instruction. Body fat content measurement was included with this version of the fitness Instruction, and a new technique for estimating body fat content was detailed. Also, new body fat standards were adopted. This figure was derived through analysis of several scientific studies by the Naval Health Research Center. The 1985 National Institutes of Health (NIH) definition of obesity has been used as an upper limit for males, with a conversion of the 1983 Metropolitan Life weight-for-height values into mean body fat percentages of 22 percent for males and 33 percent for females (Metropolitan Life Insurance Company, 1983). These figures were recommended as U.S. Navy maximums for body fat (National Institutes of Health, 1985). The recommendation for men was accepted, but command concerns about physical appearance dropped

the female standard to 30 percent. The U.S. Navy's equations were developed by Hodgdon and Beckett (1984a, b) at the Naval Health Research Center. Circumferences used for males include that of the abdomen and neck, and for females, neck, waist, and hip.

The 1990 Naval Instruction, 6110.1D (Chief of Naval Operations, 1990) made participation in the PRT for personnel more than 50 years old optional. Body fat standards were modified so that the limit for males was 22 percent and the limit for females was 30 percent. The weight-for-height table was introduced as an initial screening device for body fat evaluation.

In 1998, the female standard was unexpectedly raised back to the originally recommended 33 percent. The next Instruction, 6110.1E (Chief of Naval Operations, 1998), was released in March 1998 and contained revised weight-for-height tables and a body fat limit of 33 percent for women. Minor changes were made to the PRT testing protocol. Reports by the Government Accounting Office (1998a, b) and the Institute of Medicine (1998) recommended changes to the Military Services' fitness plans, and these suggestions led to the most recent instruction, 6110.1F (Chief of Naval Operations, 2000), the basis of the U.S. Navy's current fitness program.

## **Current U.S. Navy Fitness Program**

**Physical Fitness**—The U.S. Navy's new fitness program is governed by Naval Operations Instruction 6110.1F (Chief of Naval Operations, 2000) and consists of a biannual Physical Fitness Assessment (PFA). The PFA is a goal-oriented, total health, physical fitness and readiness program that has three components—Physical Activity Risk Factor Screening, body composition assessment, and the Physical Readiness Test (PRT) (Chief of Naval Operations, 2000). The PRT consists of the sit-reach, push-ups, curl-ups, and run or swim. This new instruction changes the U.S. Navy fitness test from a score-based to a goal-oriented program, with separate standards for each gender within age groups. All personnel are tested, with no upper age limit. There are five Performance Categories (Outstanding, Excellent, Good, Satisfactory, and Unsatisfactory) and three performance levels within these categories (High, Medium, and Low for Outstanding, Excellent, and Good Categories; High, Medium, and Marginal for the Satisfactory Category). Standards were determined on the basis of PRT scores gathered during 1997 and 1998.

The new Instruction encourages the Morale, Welfare, and Recreation (MWR) fitness staff to provide one-on-one exercise prescriptions to individuals needing assistance in attaining and maintaining their fitness level as part of the U.S. Navy's PFA.

In accordance with DoD guidance, the U.S. Navy also maintains body composition standards in addition to its physical fitness and weight requirements. Males between the ages of 17 and 39 can have a maximum of 22 percent body fat, while females in this age group can have a maximum 33 percent body fat. For personnel aged 40 and over, the limits are 23 percent (males) and 34 percent (females) (Chief of Naval Operations, 2000).

**Occupational Fitness**—In the early 1950s, the U.S. Navy started a project whose goal was to determine minimum physical specifications for U.S. Navy jobs (Browne & Germain, 1952). A literature search was conducted, and a data-gathering methodology was established. Demands of several career fields were documented. In addition, Robertson (1992) published the results of a U.S. Navy

project for developing job performance standards of muscularly demanding aircraft and shipboard tasks. The project was completed in five phases. All muscularly demanding aircraft carrier and shipboard tasks were identified, criterion tasks and performance standards were determined, an instrument was developed to measure criterion task performance, a strength test battery was constructed, and the strength test battery was validated regarding its ability to predict criterion task performance. Robertson (1992) notes that, “Although these methods to set standards have been demonstrated, they have not been implemented. Perhaps it is because (1) costs of injuries or non-performance have not been adequately demonstrated, or (2) competing concerns for selection solely on technical abilities have predominated” (p. 1301).

The U.S. Navy considered using a strength test as a screen for applicants desiring entry into physically demanding Military fields and concluded that more women than men would be excluded as a result (Government Accounting Office, 1998). Since U.S. Navy women were meeting the demands of their occupations, it was not deemed necessary to either implement such a test or categorize career fields by physical demand. While the U.S. Navy has no specific fitness standards for various occupations, it has done considerable research in this field. A study by Vickers, Hervig, and White (1997) examined the relationship between Physical Demand Ratings (PDRs) and back injury hospitalization rates (BIRs) for 73 entry-level U.S. Navy occupations. The study demonstrated a strong relationship between PDRs and BIRs. Applying the resulting quadratic function to define an exceptionally demanding job, 44 percent of the 73 occupations studied would require occupation-specific fitness standards.

Job-relevant training programs are being developed by the Naval Health Research Center in response to the DoD directive, 1308.1 (U.S. Department of Defense, 1981), which ordered each Military Service to develop training programs to meet the specific task requirements of their personnel. One such total body fitness program is SPARTEN (Scientific Program of Aerobic and Resistance Training Exercise in the U.S. Navy). SPARTEN is an on-ship program based on research findings that indicate that aerobic and circuit training is superior to aerobic and calisthenic conditioning for developing total body fitness. It offers aerobic training to maintain health and progressive resistance training, which optimizes job performance and minimizes job-related injuries (Marcinik, 1984). The movements attempted to simulate efforts such as lifting, pushing, and pulling, which are performed during the performance of muscularly demanding shipboard work. The circuit weight training exercises are performed on a multistation machine and develop all major muscle groups with minimal rest periods to further develop cardiovascular endurance.

## **History of U.S. Marine Corps Fitness Program**

Physical Fitness—U.S. Marine Corps male standards were based on 1967 studies that established 10<sup>th</sup> percentile and 90<sup>th</sup> percentile times for the 3-mile run. The 10<sup>th</sup> percentile was deemed the cut-point for failure and the 90<sup>th</sup> percentile is the upper limit for maximum points awarded. In 1997, the Corps increased the run distance for females from 1.5 to 3 miles to match the requirement for males. The female time standard was based on studies conducted from 1993 and 1996 (Government Accounting Office, 1996) establishing about a 3-minute increase in time to complete the run between males and females.

The U.S. Marine Corps Order 6100.3J (U.S. Department of the Navy, 1988) states that all U.S. Marines will participate in a minimum of 3 hours of physical fitness training per week. A successful Physical Conditioning Program is said to consist of the following types of exercises: anaerobic conditioning, progressive resistance training, and aerobic conditioning.

With respect to the development of male or female body fat standards, the U.S. Marine Corps had no available documentation (GAO, 1998a). The GAO's interview with U.S. Marine Corps officials (GAO, 1998a) revealed that the standards were based on command judgments regarding fitness and appearance, as opposed to actuarial tables or any other scientific basis. Some limited research may have been applied, however, because regulation defined the maximum allowable body fat percentage for males as 18 percent, which is just below the midpoint of the interval between the 10 percent figure said to be the average for marathon runners and the 30 percent figure that defines gross obesity. The female standard of 26 percent is at about the 80 percent point of the interval between the 11 percent body fat level which the regulation says is that of the average female gymnast and the 30 percent level that defines gross obesity in women (GAO, 1998a).

**Current Program** — The U.S. Marine Physical Fitness Program is governed by U.S. Marine Corps Order 6100.3J (U.S. Department of the Navy, 1988). The U.S. Marine Corps Physical Conditioning Program is designed to promote everyday work effectiveness, combat readiness, leadership, and self-discipline. U.S. Marine Corps testing, which is administered twice a year, differs slightly for men and women. Women and men run for 3 miles. Both sexes perform sit-ups. Women do a flexed arm hang while men perform pull-ups/chin-ups. The run is used to measure the efficiency of the cardiovascular system, and the other events are designed to test the strength and stamina of the upper body (shoulder girdle), midsection, and lower body.

Body fat standards are 18 percent (males) and 26 percent (females). Those exceeding height/weight standards are referred for body fat assessment. To determine body fat, the U.S. Marine Corps now uses a circumference equation that was first developed by Wright, Dotson, and Davis (1980, 1981) and later modified by Hodgdon and Beckett (1984a, b) which takes into account neck and abdomen measurements for men and neck, waist, and hips for women. It is based on a four-component body composition criterion. All U.S. Marines under the age of 46 participate in this testing.

**Occupational Fitness** — The U.S. Marine Corps at one time administered a physical readiness test of combat skills, but this has been discontinued.

## Evaluation of Military Fitness Standards

The previous section detailed the fitness and occupational standards of the U.S. Military Services. How these standards were determined was also documented, when this information was available. The current section reiterates the processes by which these standards were generated to underscore the limitations of these processes.

Most current Military standards were derived through the use of population norms, based on the distribution of scores obtained on a performance test. For instance, in some cases the Services have used data from Military populations, including males and females, on tests of sit-ups, push-

ups, and running to establish minimum and maximum standards at percentiles of that performance curve. In some cases, an approach this systematic was not applied to standards for females.

**U.S. Army**—Up until the time of its most current program change, the U.S. Army established minimum requirements for its U.S. Army Fitness Test on data collected in the 1980s (GAO, 1998b). These data formed the minimum requirements on which standards were based. Incremental steps to the maximums were based on simple numerical progressions from these minimums, not actual scores. Vogel (1986) reports that the U.S. Army's 2-mile timed run is a good estimate of aerobic fitness, but push-ups and sit-ups are somewhat inadequate as measures of general strength. The report also states that both of these events should be considered primarily muscle endurance measures that are limited to shoulder and abdominal muscles. While neither of these tests correlate well with common soldiering tasks, they serve to stimulate physical training activity.

Last year, the U.S. Army began to implement new standards that are based on a more statistically valid sample base than in the past. The policy behind the new standards is a gender-neutral "equal points for equal work," with minimum requirements generally set on the 8<sup>th</sup> percentile of actual scores gathered during an U.S. Army study in 1995. Maximums are based on 90<sup>th</sup> percentile scores, and requirements are reduced in five-year age increments.

**U.S. ~~Air~~ Force**—A 1998 GAO report (GAO, 1998b) concluded that U.S. Air Force officials had no published studies or other records to document the rationale for their cardiovascular endurance standards, but an informal account of the U.S. Air Force's fitness history reveals that the cardiovascular standard was based on limited normative statistics from a population of U.S. Air Force men and women in the early 1990s. The population was divided into quintiles. The GAO report states that—

*Researchers recommended that the minimum standard be set at the 20<sup>th</sup> percentile of performance because that was the point with the largest incremental gain in health benefits between percentile groups. However, U.S. ~~Air~~ Force officials wanted a higher standard for readiness reasons; as a result, the next percentile grouping, the 40<sup>th</sup> percentile, was selected as the minimum standard. Female standards were set the same way and at the same level.*

The U.S. Air Force is fielding its new strength test, with standards that will likely be based on U.S. Army standards with some possible refinements derived during this fielding period.

**U.S. Navy**—U.S. Navy standards for fitness test events: 1½ mile run/walk, push-ups, and sit-ups for men 30 and older are based on distributions of actual scores among the extant population gathered during the past two years. Earlier minimum requirements (GAO, 1998a) were set at the 10<sup>th</sup> percentile and maximums at the 90 to 95<sup>th</sup> percentiles. However, for the run time for women, an arbitrary increment of time was added to the men's standard rather than being based on actual run times of women.

**U.S. Marine Corps**—U.S. Marine Corps standards were probably based on 1967 studies showing average 3-mile run times, with maximum times set at the 90<sup>th</sup> percentile and minimums at the 10<sup>th</sup> percentile. Studies conducted in 1993 and 1996 revealed about a 3-minute difference in run times



between men and women, so this 3-minute difference was added to the men's standard scores to form the standards for females (GAO, 1998a).

## Evaluation of Military Body Composition Standards

The body composition standards set by the Department of Defense were first documented in 1981 (GAO, 1998a). A study panel made the recommendation that female and male standards be based on scientific texts findings that average body fat for physically fit males was 20 percent and average body fat for physically fit females was 30 percent. However, the actual guidance indicated a 26 percent figure for females, following the belief that, "it was desirable to recruit women whose body fat was closer to that of the average man, as such women, possessing a higher than average proportion of fat free mass, might also be more similar to men in strength and endurance." DoD standards were modified in 1995 (U.S. Department of Defense, 1995), to between 18 and 26 percent for men and 26 to 36 percent for women. GAO authors indicate that this change was made to accommodate the range of values in effect at the time from all the Services, but that no scientific research was conducted on which to base such a change.

**U.S. Army**—Friedl's 1992 chapter in *Body Composition and Physical Performance* outlines the history and rationale behind the U.S. Army's body composition standards. The U.S. Army's current 20 percent figure for males is based on actual scores of young U.S. Army males recorded during the 1980s. The 26 percent figure was attained by increasing the 20 percent figure by *two* points for every 10 years of increasing age. The U.S. Army's standards for females were determined by adding 8 percentage points to the male standard for each category. In 1991, the female standards were made less stringent (from 28 to 34 percent to 30 to 36 percent).

**U.S. Air Force**—The 1998 GAO report (Government Accounting Office, 1998a,b) indicates that U.S. Air Force officials did not have data to support their standards derivation process. The U.S. Air Force is currently considering a two-tier approach to body composition standards. The first tier would deal with health and readiness, and the second tier would represent job specific standards (Wilkinson, Kampert, Blair, Baumgartner & Constable, 2000).

**U.S. Navy**—The U.S. Navy body composition standards are based on the National Institutes of Health definition of obesity. U.S. Navy scientists converted the weight-for-height table data into mean body fat percentages of about 22 percent for males and 33 percent for females.

**U.S. Marine Corps**—The U.S. Marine Corps body fat standards appear to be based on command judgments for fitness and appearance, according to the GAO (GAO, 1998a). Some limited research may have been applied. The maximum allowable body fat for male U.S. Marines is only 18 percent, and the female standard is 26 percent.

## Evaluation of Occupational Fitness Standards

The current use of an occupational standards approach by the U.S. Services is meager. Presently, the U.S. Air Force AFSCs are categorized into eight physical demand categories, and the U.S. Air Force uses a strength aptitude test to screen out those recruits who would not be likely to perform successfully on a given job (U.S. Department of the Air Force, 1994). The U.S. Air Force does not incorporate this fitness test into the required annual fitness evaluation. The U.S. Navy has not adopted occupational strength standards for active duty personnel or recruits, nor has the U.S. Army. The U.S. Marine Corps at one time administered a physical readiness test of combat skills, but this has been discontinued.

The U.S. Air Force administers a strength test at the time of enlistment, which is used to determine job qualification. The test is an incremental lift test, called the SAT, described earlier in this chapter. At one time, the U.S. Army used a similar test, but has discontinued its use. Neither the U.S. Navy nor the U.S. Marines performs a strength assessment at any time that is occupationally focused.

---

## Summary

Interest in physical fitness and exercise is found in writings as early as 200 B.C. Citizens of ancient Greece valued those who displayed exceptional athletic performance as possessing both spiritual and physical strength rivaling the gods (USHHS, 1996). Military fitness and physical performance, or the lack of it, has long been an issue for our Military. The primary intent of physical standards in the Military has always been the selection of Service members best suited to the inherent physical job demands (Friedl, 1992). Likewise, industry showed an early interest in performance on the job and particularly the measurement of work (Hogan, 1991). For example, Taylor attempted to quantify a day's worth of work for a "firstclass man" in the late 1800s (Taylor, 1923). Much of this chapter focused on the approaches to fitness in the U.S. Military because of the easily available documentation and the stated strong interest by the Services. Furthermore, DoD guidance stipulates physical fitness programs for each of the Services. Past programs have been reviewed here in some detail. We surmise that these programs have tended to focus on "convenient" test batteries (running, calisthenics, etc.), and perhaps have had a historical bias on physical appearance. Unfortunately, the testing components rarely seem to reflect our best understanding of the science (versus the historical precedents). Previously, for example, Robertson (1992) notes that although better scientific methodologies may have been demonstrated, these improved practices have eluded programmatic implementation. Not surprisingly, a recent report by the GAO advises that the DoD should establish a mechanism for providing policy and research coordination of the Military Services physical fitness and body fat programs (DoD Joint Technology Coordinating Group-5; U.S. Army Research and Materiel Command, 1999). On the other hand, broader implementation for fitness/performance testing and standards in other governmental or commercial sectors has likely been severely constrained by potential litigation concerning perceived worker discrimination

(see Chapter 7). Other chapters in this SOAR will describe in detail the specific findings, issues, and practices associated with the development of physical fitness/performance standards to date.

---

## References

1. American College of Sports Medicine (ACSM). (1995). *Guidelines for Exercise Testing and Prescription*. Baltimore, MD: Williams & Wilkins.
2. Balke, B., & Ray, W. (1959). The present status of physical fitness in the Air Force. Unpublished research, Report No. 59-67, School of Aviation Medicine, Randolph AFB TX.
3. Barach, J.H. (1919). The energy index (S.D.R.) of the cardiovascular system. *Archives of Internal Medicine*, 24(5), 509.
4. Beumen, G. & Borms, J. (1990). Kinanthropometry: Roots, development and future. *Journal of Sports Sciences*, 8, 1-15.
5. Brouha, L. (1943). The step test: A simple method of measuring physical fitness for muscular work in men. *Research Quarterly*, 14, 31-36.
6. Browne, C. G., & Germain, G. L. (1952). *Final report on research on methods for establishing minimum physical specifications for Navy jobs*. Wayne University, Personal Research Center.
7. Bucher, C. A. (1968). *Foundations of physical education*. St. Louis, MO: Mosby.
8. Chahal, P., Lee, S. W., Oseen, M., Singh, M., & Wheeler, G. (1992). Physical fitness and work performance standards: A proposed approach. *International Journal of Industrial Ergonomics*, 9, 127-135.
9. Chief of Naval Operations. (1976). Physical fitness. Office of the Chief of Naval Operations Instruction 6110.1. Washington, DC: Office of the Chief of Naval Operations. 16 June.
10. Chief of Naval Operations (OP-09). (1980). Physical fitness. Office of the Chief of Naval Operations Instruction 6110.1A. Washington, DC: Office of the Chief of Naval Operations. 17 July.
11. Chief of Naval Operations (NMPC-6H). (1982). Health and physical readiness program. Office of the Chief of Naval Operations Instruction 6110.1B. Washington, DC: Navy Military Personnel Command. 19 October.
12. Chief of Naval Operations (NMPC-68). (1986). Physical readiness program. Office of the Chief of Naval Operations Instruction 6110.1C. Washington, DC: Navy Military Personnel Command. 7 August.
13. Chief of Naval Operations (NMPC-68). (1990). Physical readiness program. Office of the Chief of Naval Operations Instruction 6110.1D. Washington, DC: Navy Military Personnel Command. 18 January.
14. Chief of Naval Operations (PERS-601). (1998). Physical readiness program. Office of the Chief of Naval Operations Instruction 6110.1E. Washington, DC: Navy Military Command. 23 March.
15. Chief of Naval Operations. (2000). Physical readiness program. Office of the Chief of Naval Operations Instruction 6110.1F. Washington, DC: Office of the Chief of Naval Operations. 1 May.
16. DiDonato, J. (2000). Fitness through the ages. *Master fitness trainer course*. U.S. Army Fitness School, Fort Benning.

17. DoD Joint Technology Coordination Group-5 & U.S. Army Medical Research & Material Command. (1999). Research Workshop on Physical Fitness Standards and Measurements within the Military Services. Summary Report.
18. Fleishman, E. A. (1964). *Structure and measurement of physical fitness*. Englewood Cliffs, NJ: Prentice Hall.
19. Foster, W. L. (1914). A test of physical efficiency. *American Physical Education Review*, 19, 632.
20. Friedl, K. E. (1992). Body composition and military performance: Origins of the Army Standards. In B.M. Marriott, & J. Grumstrup-Scott (Eds.), *Body Composition and physical performance applications for the military services*, 31–55. Committee on Military Nutrition Research Food and Nutrition board, Institute of Medicine. Washington, DC: National Academy Press.
21. Government Accounting Office. (1996). *Physically demanding jobs: Services have little data on ability of personnel to perform* (NSIAD-96-169), July 9. Washington, DC: Author.
22. Government Accounting Office. (1997). *Military personnel: First term recruiting and attrition continue to require focused attention* (GAO/T-NSIAD-00-102), February 24. Washington, DC.
23. Government Accounting Office. (1998a). *Gender issues — Improved guidance and oversight needed to ensure validity and equity of fitness standards*. Washington, DC: Author.
24. Government Accounting Office. (1998b). *Gender issues — Information to assess service members' perceptions & gender inequality is incomplete*. Washington, DC: Author.
25. Gilbreth, F. B. (1901). *Bricklaying system*. New York Clark.
26. Hackensmith, C. W. (1966). *History of physical education*. New York Harper & Row.
27. Hartwell, E. M. (1885). *Physical training in American colleges and universities* (Circular of Information No. 5, 59–92). Washington, DC: Bureau of Education.
28. Hitchcock, E., & Seelye, H. H. (1893). *An anthropometric manual*. Amherst, MA: Carpender & Morehouse.
29. Hodgdon, J. (1999). *A history of the U.S. navy physical readiness program from 1976 to 1999*. Arlington, VA: Office of Naval Research.
30. Hodgdon, J. A., & Beckett, M. B. (1984a). *Prediction of percent body fat for U.S. navy men from body circumferences and height* (Report No. 84-11). San Diego, CA: Naval Research Center.
31. Hodgdon, J. A., & Beckett, M. B. (1984b). *Prediction of percent body fat for U.S. navy women from body circumferences and height* (Report No. 84-29). San Diego, CA: Naval Research Center.
32. Hoeger, W. K., & Hoeger, S. A. (1998). *Lifetime physical fitness and wellness* (5<sup>th</sup> ed.). Englewood, CO: Morton Publishing Company.
33. Hogan, J. C. (1991a). *Physical abilities: Chapter 11 Handbook & industrial and organizational psychology* (2<sup>nd</sup> ed., Vol. 2). Palo Alto, CA Consulting Psychologists Press, Inc.
34. Hogan, J. (1991b). Structure of physical performance in occupational tasks. *Journal of Applied Psychology*, 76, 495–507.
35. Hunsicker, P. A. (1958). *A.A.H.P.E.R. youth fitness manual*. Washington, DC: A.A.H.P.E.R., Youth Fitness Project.
36. Institute of Medicine. (1998). *Assessing readiness in military women: The relationship of body composition, nutrition, and health*. Washington, DC: National Academy Press.
37. Kellogg, J. H. (1896). *The values of strength tests in the prescription of exercise* (Vol 2). Battle Creek, MI: Modern Medicine Library.

38. Kraus, H., & Hirshland, P. R. (1954). Minimum muscular fitness test in school children. *Research Quarterly*, 25, 177–188.
39. MacCurdy, H. L. (1933). *A test for measuring the physical capacity of secondary school boys*. New York, NY: Author.
40. Marciniak, E. J. (1984). *SPARTEN: A total body fitness program for health and physical readiness*. San Diego, CA: Naval Health Research Center.
41. Marriott, Bernadette M., & Grumstrup-Scott, Judith (Eds.). (1992). *Body composition and physical performance: Applications for the military services*. Washington, D.C.: National Academy Press.
42. Maud, P. J., & Foster, C. (1995). *Physiological assessment of human fitness*. Champaign, IL: Human Kinetics.
43. McCurdy, J. H., & Larson, L. A. (1939). *The physiology of exercise*. Philadelphia: Lea & Febiger.
44. Medical Command, Department of the Air Force. (1996). *The Air Force Fitness Program*, Air Force Instruction 40–501. Washington, DC: Author.
45. Metropolitan Life Insurance Company. (1983). Metropolitan Life Insurance Company: 1983 Metropolitan height and weight tables. *Statistical Bulletin*, 64, 2–9.
46. National Institutes of Health. (1985). Health implications of obesity. In *National Institutes of Health Consensus Development Conference Statement*, 5. Washington, DC: U.S. Department of Health and Human Services.
47. Navy News Service. (1994). New Navy health and physical readiness program release. [On-line] <http://www.chinfo.navy.mil/navpalib/news/nns94/nns9405.txt>
48. Palmer, B. & Soest, J. (1997). *Expanded Air Force physical fitness battery: Muscle strength, muscle endurance, and flexibility considered* (CSERIAC–RA–97–003A) Wright Patterson Air Force Base, OH: Crew System Ergonomics Information Analysis Center.
49. Pollock, M. L., Garzarella, L., DeHoyos, D., Brechue, W., Beekley, M., Werber, G., & Lowenthal, D.T. (1994). *The Cross-validation of the United States Air Force submaximal cycle ergometer test to estimate aerobic capacity*. Brooks AFB: Crew Systems Directorate.
50. Robertson, D. W. (1992). Development of job performance standards for muscularly demanding military tasks. In S. Kumar (Ed.), *Advances in industrial ergonomics and safety IV* (pp. 1299–1304). Washington, DC: Taylor & Francis.
51. Rogers, F. R. (1927). *Tests and measurement programs in the redirection of physical education*. New York: Columbia University Bureau of Publications.
52. Sargent, D. A. (1921). The physical test of man. *American Physical Education Review*, 25, 188–194.
53. Schellhous, R. R. (1982). *Air force physical fitness: An assessment of characteristics and programs which affect individual physical fitness*. Wright-Patterson AFB, OH: Air Force Institute of Technology (DTIC No. AD–A123 022).
54. Schneider, E. C. (1920). A cardiovascular rating as a measure of physical fitness and efficiency. *Journal of the American Medical Association*, 74 (5), 1506–1507.
55. Seaver, J. W. (1890). *Anthropometry in physical education*. New Haven, CT: Tuttle, Morehouse & Taylor.
56. Section 543 of the FY 1994 National Defense Authorization, Public Law 103–160, November 30, 1993.
57. Sharp, S. E. (1898). Individual psychology: A study in psychological methods. *American Journal of Psychology*, 10, 329–391.

58. Sheldon, W. H. Stevens, S. S. & Tucker, W. B. (1940). *The varieties of human physique*. New York: Harper and Brothers.
59. Taylor, F. W. (1923) *Principles of scientific management*. New York: Harper.
60. Thorndike, R. L. & Hagen, E. (1969) *Measurement and evaluation in psychology and education* (3<sup>rd</sup> ed.). New York: John Wiley & Sons, Inc.
61. Tuttle, W. W. (1931) The use of the pulse rate test for rating physical efficiency. *Research Quarterly*, **11** (2), 5.
62. U.S. Army Physical Fitness School. (2000). World War I. [On-line]  
<http://192.153.150.25/usapfs/pages/history1885%2D1920.htm>
63. U.S. Department of the Air Force. (1962). *The Royal Canadian Air Force 5BX plan for physical fitness*. AFP 50-5-1. Ottawa: Queen's Printer and Controller of Stationery
64. U.S. Department of the Air Force. (1947). *Training — Physical conditioning* (Air Force Regulation 50-5). November 13. Washington, DC: Government Printing Office.
65. U.S. Department of the Air Force. (1959). *Training — Physical conditioning*. (Air Force Regulation 50-5). April 27. Washington, DC: Government Printing Office.
66. U.S. Department of the Air Force. (1961). *Physical conditioning*. (Air Force Manual 160-126) June 8. Washington, DC: Government Printing Office.
67. U.S. Department of the Air Force. (1977) *Air Force physical fitness program*. (Air Force Regulation 35-11, Immediate Message Change 79-1). September 23. Washington, DC: Government Printing Office.
68. U.S. Department of the Air Force. (1981) *Air Force physical fitness program*. (Air Force Regulation 35-11) July 28. Washington, DC: Government Printing Office.
69. U.S. Department of the Air Force. (1994) *Airman classification*. (Air Force Manual 36-2108). October 31. Washington, DC: Government Printing Office.
70. U.S. Department of the Air Force. (1994). *The weight management program* (Air Force Instruction 40-502). November 7. Washington, DC: Government Printing Office.
71. U.S. Department of the Army. (1998) *Physical fitness training* (Field Manual 21-20). October 1. Washington, DC: Government Printing Office.
72. U.S. Department of Defense. (1981). *Physical fitness and weight control programs* (Department of Defense Directive 1308.1) June 29. Washington, DC: Government Printing Office.
73. U.S. Department of Defense. (1985) *Joint DoD committee on fitness references directive* (Department of Defense Directive 1308.2) January 25. Washington, DC: Government Printing Office.
74. U.S. Department of Defense. (1995) *Physical fitness and body fat programs procedures* (Department of Defense Directive 1308.3) August 30. Washington, DC: Government Printing Office.
75. U.S. Department of the Navy. (1988) *Physical fitness*. (Marine Corps Order 6100.3). February 26. Washington, DC: U.S. Marine Corps Headquarters.
76. Vickers, R. R., Hervig, L. K. & White, M. R. (1997) *Job demands and back injury in Navy personnel*. San Deigo, C A Naval Health Research Center.
77. Vogel, J. A. (1986). *Fitness and activity assessments among U.S. Army populations: Implications for NCHS general population surveys*. Natick, MA: U.S. Army Research Institute of Research Medicine.
78. Vogel, J. A., Kirkpatrick, J. W., Fitzgerald, P. I., Hodgdon, J. A. & Harman, E. A. (1988) *Derivation & anthropometry based body fat equations for the Army's weight control program*. (Technical Report No. T17-88). Natick, MA: U.S. Army Research Institute of Environmental Medicine.

79. Wilkinson, W. J., Kampert, J. B., Blair, S. N., Baumgartner, N., & Constable, S. H., (2000). Physical fitness and health: A comparative review of the USAF fitness programs. The Cooper Institute for Aerobics Research & The U.S. Air Force School of Aerospace Medicine, PE Division, USAF School of Aerospace Medicine.
80. Wissler, C. (1901). Correlation of mental and physical tests. *Psychological Monographs*, 16.
81. Wright, H. F., Dotson, C. O., & Davis, P. O. (1980). An investigation of assessment techniques for body composition of women Marines. *US Navy Medicine*, 71, 15–26.
82. Wright, H. F., Dotson, C. O., & Davis, P. O. (1981). A simple technique for measurement of percent body fat in man. *US Navy Medicine*, 72, 23–27.





## Chapter 2

# Health-Related Fitness Standards: A Baseline Approach

Stefan H. Constable, Ph.D.  
USAF School of Aerospace Medicine  
Brooks AFB, TX

---

### Abstract

This chapter attempts to build a rationale for an ancillary or alternative approach to fitness standards development: health-based fitness levels. The difficulty with instituting job-specific performance tests in the Military or elsewhere is identified throughout this State-of-the-Art Report (SOAR). Moreover, performance on current Military fitness tests does not correlate well with performance on task-specific job tests required for many Military vocations. The scientific literature is now replete in supporting the strong association between physical activity/fitness and general health, wellness, and quality of life. Therefore, we postulate that health-based standards seem to be an adjunctive approach to the physical fitness standards process. There may be greater specific application opportunities for the Military. The focus of this chapter is to—

1. Document the specific relationship between physical activity or fitness and specific health outcomes,
2. Review exercise prescriptions and investigate the quantitative relationships between physical activity benefits and measured levels of fitness,
3. Identify those attempts to produce a specific cut-point for the specified (identified) fitness components, and
4. Assess at least qualitatively, the validity of those health-based fitness approaches.

Identifying the minimal dose-response (versus adequate or optimal) relationships, not to mention truly testable metrics or standards, has proved the greatest challenge. However attractive or meritorious this endeavor may initially seem, the basic observation should be that in application or practice varying levels of difficulty may be encountered, depending on the chosen fitness modality application, that is, aerobic, strength, or body composition. This possibility is primarily due to the lack of suffi-

cient data and/or discrete methodologies to identify clearly defined cut-points on which to base standards. Nevertheless, this should not deter further efforts to investigate or apply alternative procedures.

---

## Health and Fitness Across the Workplace

Developing fitness standards has usually been determined by some minimum level of individual physical capacities that relate well to occupational job performance for candidate selection, worker retention, or advancement. Although there are currently many workplace scenarios in which the physical requirements of the job are often quite low, there is still significant interest in achieving or maintaining a reasonable fitness level for all workers. For a variety of reasons, including better health, everyday work efficiency, improved cognitive functioning, good appearance, and increased readiness, fitness is essential especially in the Military. This, of course, implies a corresponding measure or metric of some level of physical fitness, with a resultant, positive impact on occupational performance. Thus, the issue of a general or baseline fitness requirement in the workplace population or the Military is twofold—

1. A basic level of fitness for overall health, and
2. Increased levels of fitness for optimum performance for occupational and recreational activities.

Currently, the Department of Defense considers physical fitness an important component of the “general health and well-being” and readiness of all Military members (Institute of Medicine, 1998). Therefore, this chapter explores the concept of baseline, health-related fitness requirements with potential application to selected Military and/or civilian environments. This chapter also presents the theoretical merit for this more generic approach to the process of physical fitness standards development, and identifies the methodological procedures and precedents for further application.

---

## General Health and Fitness Descriptors

First, it is necessary to identify a framework or general definition associated with health. Although the specific nomenclature used may be diverse, individual health status clearly exists on a continuum as does physical fitness (Institute of Medicine, 1998). The World Health Organization has defined health as a positive state of physical, mental, and social well-being (U.S. Department of Health and Human Services, 1996) as opposed to just the absence of disease or disability. In addition, health-related quality of life (HRQL) has received considerable attention on clinical, scientific, and public-interest fronts (U.S. Department of Health and Human Services, 1996).

The preceding terminology relates well to the detailed characterizations of health-related fitness expressed in Bouchard and Shephard’s chapter in the *Second International Consensus Symposium on Physical Activity, Fitness, and Health* (1993)—

*Fitness is operationalized in present-day Western societies with a focus on two goals: performance and health. Performance-related fitness refers to those components of fitness that are necessary for optimal work or sport performance. Health-related fitness refers to those components of fitness that are affected favorably or unfavorably by habitual physical activity and relate to health status. It has been defined as a state characterized by (a) an ability to perform daily activities with vigor, and (b) demonstration of traits and capacities that are associated with a low risk of premature development of hypokinetic diseases and conditions. Important components of health-related fitness include body mass for height, body composition, subcutaneous fat distribution, abdominal visceral fat, bone density, strength and endurance of the abdominal and dorso-lumbar musculature, heart and lung function, blood pressure, maximal aerobic power and capacity glucose and insulin metabolism, blood lipid and lipoprotein profile, and the ratio of lipid to carbohydrate oxidized in a variety of situations. A favorable profile for these various factors presents a clear advantage in terms of health outcomes as assessed by morbidity and mortality statistics. (p. 15)*

For this chapter, three main categories of health-related fitness are identified as follows—

1. Cardiorespiratory fitness is the ability of the respiratory and circulatory systems to adapt to and recover from vigorous activities that involve large muscle groups, thus increasing the heart rate and blood circulation. Examples of these activities are walking, jogging, swimming, and biking.
2. Body composition is the ratio of body fat to lean body tissue (muscle and bones, etc.) expressed as a percentage.
3. Musculoskeletal fitness may be subdivided into muscular strength, muscular endurance, and whole-body flexibility.

Strength is measured by determining the one-repetition, maximal force that can be exerted against a resistance, for example, lifting a weighted bar. Muscular endurance is the ability to repeatedly apply force to resistance. Endurance, for example, could be measured by the number of repetitions one can perform of a particular exercise such as sit-ups. Flexibility is the ability to operate through a "full" range of motions such as touching one's toes while in a seated position with the legs straight (Nieman, 1998).

Often these definitions tend to discriminate physical skills such as speed, agility, power, and coordination from physical fitness. Although physical skill is useful for sports and certain job/task performance (and may improve with increased physical activity), it is important to note that athletic skill is generally not necessary for health improvement or disease prevention.

Table 2.1 (derived from Pate, 1988) demonstrates the overlap among more specific components suggested to fall under the generic fitness label as they might relate to motor performance, physical fitness, and health-related fitness.

## Metabolic Fitness

More recently, an analogous category of health-related fitness —metabolic fitness—was introduced by Despres et al. (1990) and conceptualized in the American College of Sports Medicine

**Table 2.1 Components of motor performance, physical fitness, and health-related physical fitness**  
*Journal of the American Medical Association, 1995, 273, 40.*

Component	Motor Performance	Physical Fitness	Health-related Physical Fitness
Anaerobic Power			
Speed			
Muscular Strength			
Muscular Endurance			
Cardiorespiratory Endurance			
Flexibility			
Body Composition			
Agility			

*Position Stand* (Pollock et al., 1998). This concept “describes the state of metabolic systems and variables predictive of the risk of diabetes and CV disease which can be favorably altered by increased physical activity or regular endurance exercise without the requirement of a training-related increase in  $VO_{2max}$ ” (p. 976). Gaesser (1996) approaches a desirable level of metabolic fitness in yet another way. He prescribes a healthy diet, (less than or equal to 20 percent fat intake), and moderate and frequent exercise at the least. Good metabolic fitness will be then evidenced by normal outcomes for blood pressure, blood sugar, and lipid profiles, regardless of the resultant body weight. The key here is really a low-fat diet, basically without caloric restriction, with the general focus on a healthy lifestyle without attention to body weight.

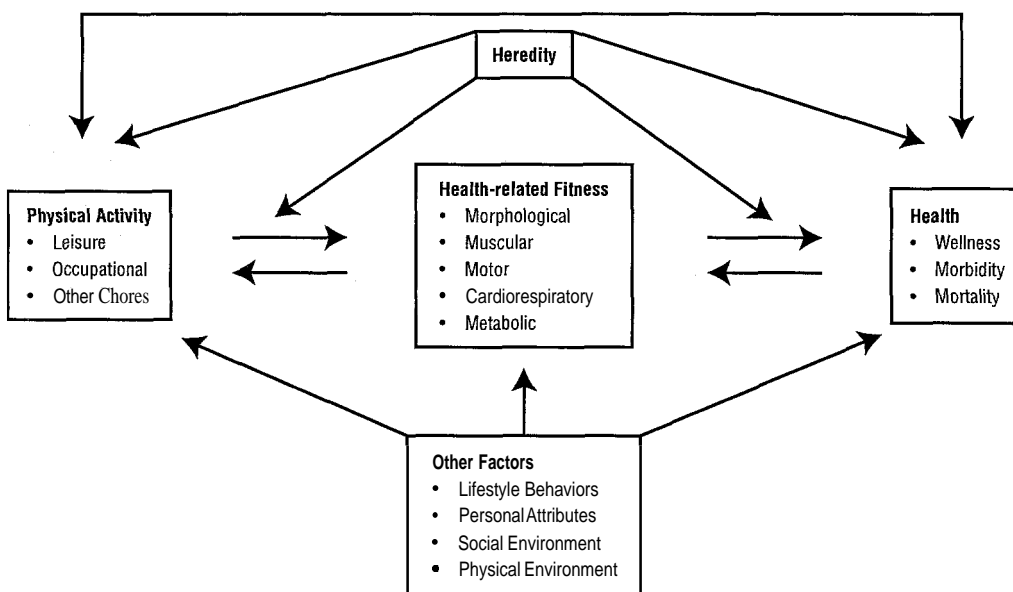
Although this metabolic approach to general fitness may have significant scientific merit, it seems most problematic if one is interested in validly monitoring compliance: abnormal medical measures/outcomes, that is, blood pressure, blood sugar, lipid profiles, and intra-abdominal fat may not be solely due to exercise and diet habits. In other words, how would one evaluate the true rigor with which each person followed the exercise and diet prescription other than by self-report? It certainly would not be feasible in a highly accountable Military setting. Moreover, these concerns are especially valid for the younger population in which the metabolic consequences of noncompliance can take years to manifest into morbidities or premorbidities.

## The Relationships Between Physical Activity, Fitness, and Health

Although improving one’s physical performance is desired by many people, maintaining or improving one’s personal health is by far the number one priority. This is especially true in today’s increasingly health-conscious environment, in which the physical requirements of our jobs are consistently diminishing. Improvements in fitness levels generally correlate very well with both increased physical performance and health. Clearly, the greatest health benefits are derived when one improves from low or modest levels of initial fitness (Blair, 1995). On the other hand, elite athletes must expend significant amounts of time and effort to achieve relatively very small increases

in performance improvement (or fitness). Ironically, this latter goal may even be at the expense of good health because of increased injuries, risk of infection, and other physiological maladies. Still, the relationships between physical activity, fitness, and health tend to be strongly and positively related: more activity is generally better. It is noteworthy that these parameters also tend to reside on a continuum of effects with a distinct degree of interindividual variability.

Although it has been argued that the degree of improvement in health status is often closely dependent on the magnitude of the improvement in fitness, it is becoming more apparent that this relationship is not so simple (Haskell, 1994). Indeed, Bouchard and Shephard (1993) state, "...the relationships between the levels of physical activity, health-related fitness, and health are complex" (p. 11). These scholars generally define health-related fitness as one's ability to perform daily activities with vigor (Institute of Medicine, 1998) and have attempted, at least notionally, to describe these associations in a model of health-related fitness as shown in Figure 2.1. This model clearly supports our general concept of an ancillary approach to fitness standards development: health-based fitness levels.



**Figure 2.1 Relationship among habitual physical activity, health-related fitness, and health status.** Reprinted, by permission, from C. Bouchard, R. J. Shepard, & T. Stephens, 1993. *Physical Activity, Fitness, and Health: Consensus Statement*. Champaign, IL: Human Kinetics, 12.

Activity and fitness levels are therefore associated in a reciprocal manner: If a person engages in more activity, fitness level improves. Conversely, if a person incorporates little or no activity in his or her life, fitness level diminishes. In addition, a higher level of fitness increases one's ability to engage in a broader variety of activities and at greater intensities that in turn provides a greater level of fitness, and therefore more opportunity to become even fitter. In this case, the self-fulfilling prophecy is a positive one. At the initial onset of activity, there may be minimal physical ability. By repeating the exercises one can adapt, improving physical health, enabling more activities to be

undertaken, and ultimately improving readiness. The close association and tightly interlaced nature of exercise and health are vividly apparent throughout the research regarding these two elements.

One of the major recognized avenues for maintaining or improving health is through physical activity. The Surgeon General's report (U.S. Department of Health and Human Services, 1996) has defined physical activity as "bodily movement produced by the contraction of skeletal muscle that increases energy expenditure above the basal level" (p. 20). The American College of Sports Medicine *Position Stand* (1990) further indicates that the adaptive responses to physical training (or exercise) are complex and normally include peripheral, central, structural, and functional factors (Pollock et al., 1998). Individual, physiologic adaptations have generally been well characterized in the literature. However, comparable data are insufficient relative to the specific intensity, frequency, and duration of training necessary to confidently predict the precise quantification of the benefit outcomes (Pollock et al., 1998). In other words, it is very difficult to specifically characterize each of the individual dose-response relationships.

In the Physical Activity, Fitness, and Health International Proceedings and Consensus Statement (Bouchard, Shephard, & Stevens, 1994), Haskell offers a more detailed description of the physiological effects of exercise (defined as any sustained or repeated movement of a relatively large skeletal muscle mass) that occur through the body's immediate response to the exercise, the adaptations that occur over time as the body increases its capacity or efficiency, or some combination of these acute and chronic responses. Haskell further explains—

*During and following this muscle contraction, local biochemical factors, along with activation of the central nervous system, stimulate the increased activity of various hormones and enzymes that help regulate key metabolic functions; and there are major shifts in cardiorespiratory performance. If the activity is of sufficient intensity and duration, the renal, hepatic, gastrointestinal, and immune systems become involved. Also, physical activity exerts physical forces on the bones, muscles, and connective tissue as a result of muscle contraction or in response to gravity. (p. 1030)*

---

## Specific Impact of Activity and Fitness on Personal Health

"Throughout history, numerous health professionals have observed that sedentary people appear to suffer from more maladies than active people." This early example may be found in the writings of English physician Thomas Cogan, author of *The Haven of Health* (1584); he recommended his book to students who, because of their sedentary ways, were believed to be most susceptible to sickness (U.S. Department of Health and Human Services, 1996).

The influence of physical activity on disease prevention and mitigation, longevity, and overall health has been under continual scrutiny for the past few decades. Several intensive reviews of these topics have been mentioned earlier, including *Physical Activity and Health — A Report of the Surgeon General* (U.S. Department of Health and Human Services, 1996), and *Physical Activity, Fitness and Health Consensus Statement* (Bouchard, Shephard, & Stevens, 1993). Other studies by leading experts have scientifically explored the beneficial effects of physical activity on several major conditions and diseases. For example, these afflictions include obesity (DePietro, 1995; Stefanick, 1993), hyperten-

sion (Hagberg, Montain, Martin, & Ehsani, 1989), cardiovascular disease (CVD) (Lee, Blair, & Jackson, 1999; Stofan, DiPietro, Davis, Kohl, & Blair 1998; Farrell et al., 1998), and diabetes (Hu et al., 1999). Other studies have investigated the possible benefits of physical exercise on osteoporosis (Recker et al., 1992), cancer (Sesso, Paffenbarger, Ha, & Lee, 1998), clinical depression, stroke (Kiely, Wolf, Cupples, Beiser, & Kannel, 1994), and musculoskeletal health (e.g., back pain, limb function). Refer to Appendix A for more detail.

Moreover, physical activity has been shown to have salutary effects on more than one condition at a time. For example, frequent exercise has been shown to alleviate obesity (DePietro, 1995), and obesity has been associated with other conditions, including diabetes (Helmrich, Ragland, Leung, & Paffenbarger, 1991), hypertension (Wier, 1992), CVD (Lee, Blair, & Jackson, 1999), and stroke (Gorelick et al., 1999). In addition to alleviating obesity as a risk factor for other conditions, physical activity has been studied in direct relation to the prevention of diseases and reduction in overall mortality rates (Lee, Blair, & Jackson, 1999; Blair et al., 1989).

## **Interrelated Impact on Conditions**

The broad implications of engaging in physical activity create a positive domino effect on disease prevention. In other words, moderate levels of physical activity can have very positive benefits for many different illnesses and conditions at the same time. For example, if a person can control obesity by exercising regularly, other disease risks are reduced for such illnesses as diabetes, heart disease, and even clinical depression. By eliminating (or reducing the severity of) obesity through physical activity, the risk of hypertension diminishes. This is also true of stroke whose risk can be reduced twofold because hypertension and obesity, two of the biggest risk factors for stroke onset, can be controlled simultaneously. In addition, the risk of diabetes, coronary heart disease, cancer, and possibly even depression is reduced. As each condition is prevented, a person's overall health, longevity, and quality of life is improved. The overwhelming benefit is that all of these risks may be reduced by the same intervention—moderate physical activity that is part of one's everyday habits. Moreover the feeling of accomplishment and improved well-being further promotes greater exercise compliance.

This proposition is supported by research findings. Lee, Blair, and Jackson (1999) examined the health benefits of leanness in relation to cardiorespiratory fitness and all-cause mortality. They collected data on 21,925 men, aged 30 to 83. Over the course of eight follow-up years there were 428 deaths. The researchers adjusted for age, examination year, cigarette smoking, alcohol intake, and parental history of ischemic heart disease. They found that the unfit lean men had twice the risk of all-cause mortality than that of fit lean men as measured by maximal exercise testing and body composition assessment. They also discovered that unfit, lean men had a higher risk of all-cause and CVD mortality than the men who were fit and overweight. Similarly, incorporating waist girths as a measure of fatness, the men who were physically unfit had a higher risk of all-cause mortality than those who were fit, even if the unfit men were thinner. Those with a greater fitness level, regardless of fat mass, tended to be at lesser risk. The authors concluded that the health benefits of leanness are limited to fit men, and being fit may reduce the hazards of overweight and obesity.

A further example of the interrelatedness of conditions is found in the connection between depression and coronary heart disease. According to the National Institute of Mental Health

(NIMH) (1999), “Research over the past two decades has shown that depression and heart disease are common companions and what is worse, each can lead to the other.” The findings regarding depression preceding heart disease show a more rapid heartbeat, higher blood pressure, and faster blood clotting in those who are depressed. Elevated insulin and high cholesterol may also be side effects of depression and are contributors to heart disease. Other NIMH studies found that depressed, heart disease-free patients were 4 times more likely to have a heart attack in the subsequent 14 years. In addition, these studies indicated that heart patients who were depressed were four times more likely to die in the subsequent six months than those who were not depressed. Although one in six people will have an episode of major depression in his or her lifetime, the number increases to one in *two* for those who have heart disease. Since depression is the leading cause of disability worldwide and heart disease is the leading cause of death in the United States, the combination of these conditions is synergistic and results in a major impact on individual health and our society’s economic growth.

Furthermore, exercise is specified as an effective intervention for both heart disease and depression and also for other conditions (e.g., hypertension, insulin sensitivity) associated with those two diseases. Hypertension cannot be ignored when discussing the relationship between diseases. It is a major risk factor for several conditions (e.g., stroke, cardiovascular disease, and diabetes). The effects of physical activity on reducing blood pressure are clearly documented and directly related. The current recommendation to engage in moderate physical activity most days of the week has been shown to have a salutary effect on reducing hypertension. The key factor for reducing blood pressure is continual, frequent activity throughout life. It is possible that if a few key risk factors (e.g., obesity, hypertension, and depression) for specifically life-threatening illnesses (e.g., heart disease or stroke) are reduced, then the rate of multiple disease conditions will be reduced as well. The consistency of the research findings lies in the high positive correlation between frequent activity at any level and salutary effects on so many of the conditions described in this chapter and Appendix A. With all of the evidence relating the aforementioned diseases to an inactive, sedentary lifestyle, there is a lot of reason to heed the government’s standards of regular, moderate activity.

## Reduced/Increased Risk of Injury

Workers who are fit are less likely to have on-the-job injuries. An extensive review of the role of physical training in preventing occupational injuries can be found in a 1992 *Ergonomics* article entitled, “Physical Training: A Tool of Increasing Work Tolerance Limits of Employees Engaged in Manual Handling Tasks” (Genaidy et al., 1992). The article emphasizes that a lack of physical fitness is a contributing factor to musculoskeletal injuries resulting from manual material handling in particular (Palmer & Soest, 1997).

A recent supplement to the *American Journal of Preventive Medicine*, entitled, *Injuries in the U.S. Armed Forces: Surveillance, Research and Prevention* (Jones & Amoroso, 2000) has focused on the specific issue of injuries in the Military. Altarac et al. (2000) point out that among the number of **risk** factors for injury previously identified, low physical fitness and lower amounts of physical activity are perhaps in the top five. Some risk factors may tend to be additive in their effects as well. Not surprisingly, prior smoking tended to predict lower levels of physical performance in U.S. Army recruits.



Obviously, exercise produces extensive benefits, but there are also possible risks associated with regular physical activity. Although the most severe of these risks may be the risk of heart attack, the most prevalent and critical risk of an exercise program would be to not continue it. A sedentary lifestyle with little or no sustained physical activity is the precursor to being physically unfit, which in turn has been demonstrated to be associated with increased risk of illness and decreased quality of life. Therefore, to become fit, a lifestyle change must occur. When implementing changes in activity level, precautions must be taken to avoid musculoskeletal injuries that would obviously deter people from further activity. To protect oneself against the risks associated with beginning an exercise program, a physically unfit person should check with his or her doctor for any existing conditions that would require special attention (e.g., extremely high blood pressure, existing heart disease, or musculoskeletal problems). The precautions set by the physician should be followed, the activity program should be pursued in a gradual and rational manner, and the activity itself should be modified to each person's beginning fitness level.

To become reasonably fit, one must frequently engage in low- to moderately-intense physical activity. As mentioned previously, understanding that highly intense, vigorous, and prolonged exercise is not required to improve health is essential in moving from no activity to some activity. Such vigorous exercise is intimidating and risky for a physically unfit person. When first starting a program, performing light exercise reduces the risks of developing strained, pulled, or overly sore muscles. When people exercise in excess of what their bodies can handle, they can incur injuries. Researchers have found a greater risk of injury associated with those having both higher- and lower-than average body mass index (BMI) (Jones et al., 1992). Since many sedentary people fall outside the average BMI category, it is important to note the possibility of injury and the importance of proper exercise techniques (e.g., warming up, stretching, and cooling down). In addition, knowing the beneficial effects of light exercise or even increasing everyday activities (such as walking up and down stairs) is encouraging for those who are not ready or willing to engage in a program of vigorous activity. The risk of injury or even discomfort, which may preclude regular exercise habits, is diminished by the current "increased lifestyle activity" recommendations. The implication of this approach is that more people will move from the "physically unfit" to the "moderately fit" category, thus significantly reducing mortality and morbidity rates.

In a study of 100 young athletes (average age 18) who died suddenly while exercising, 90 had heart or blood vessel birth defects. Congenital CVD is now the major cause of athletic death in high school and college (Nieman, 1998). Other reasons for sudden heart attack while exercising are known or unknown heart disease, usually related to a family history of the disease. A study by researchers at Harvard found heart-attack risk to be 5.9 times higher after heavy physical exertion versus lighter or no exertion (Nieman, 1998). This risk was especially acute for those who were habitually inactive, while those who were accustomed to exercise had much less risk of heart attack when engaging in heavy activity. An additional study of 36 marathon runners who had heart attacks found that most of the runners had a strong family history of heart disease, suffered early warning symptoms of the disease itself, but did not take precautions. Most researchers have found that fewer than 10 out of 100,000 men will have a heart attack during exercise. Most likely these were men who were sedentary and then engaged in physical activity with at least a high risk for heart disease. For this reason, it is important for people to be screened and cleared by their physicians before starting an activity program. The ACSM (1994) states that "The incidence of cardiovascular problems during physical activ-

ity is reduced by nearly 50 percent when individuals are first screened, and those who are identified with risk factors are diverted to other professionally established activity programs” (pp. i–v).

Although there are risks associated with engaging in activity, the greatest risk is engaging in no activity. People at all fitness levels can safely implement the current activity recommendations for good health, tailoring a program to specific needs if necessary. Clearly, if a person takes the aforementioned precautions and consistently engages in physical activity throughout life, the death rate from preventable diseases would drop significantly.

## **General Sense of Well-Being**

In addition to the mental health benefit of less depression alluded to earlier, regular exercise can also reduce levels of anxiety, tension, and reaction to life’s stresses. Unfortunately, a person’s improved well-being is difficult to assess and quantify. In addition, an individualized program that will assuredly affect psychological well-being is difficult to prescribe. Still it is clear that physical activity can positively affect many areas of mental health (Bouchard et al., 1993).

Kaplan and Bush (U.S. Department of Health and Human Services, 1996) perhaps took this further in a construct representing a person’s overall satisfaction with health-related quality of life (HRQL). The intent was to capture the influence that health status and care have on the quality of life. Specific effects of exercise on HRQL include psychological well-being, perceived physical function and well-being, and perhaps cognitive function. A recent review (McAuley, 1994) suggested that positive associations between self-esteem and physical activity exist for both young adults and children. This was also true whether the activity was chronic (long-term training) or acute (a single bout of activity). These findings are basically independent of age, and with older adults significant improvements in psychological well-being may be found without improvement in aerobic fitness.

## **Everyday Work Performance and Cognition**

Besides deployment readiness, the everyday work environment benefits from a fit work force. Workers who are fit are more productive, happier, absent less often, and incur fewer on-the-job injuries. Healthy and fit workers are more productive. They suffer less from fatigue and make fewer errors (Shephard, 1992). Fit workers file fewer insurance claims, and injure themselves less frequently. A NASA and U.S. Public Health Service survey (Durbeck et al., 1973) of more than 200 Federal employees who participated in a worksite exercise program revealed that workers who exercised felt that, as a result, they could work harder mentally and physically, they enjoyed their work more, and found their normal work routine less boring.

Overall, a number of studies tend to suggest potentially beneficial associations between worker health and productivity. Although the relationships are statistically significant, the true strength of the relationships may be generally low (Bouchard et al., 1993). Therefore uncertainty exists regarding the expected cost-benefit ratio when instituting an employee fitness program. In fact, a review of numerous health promotion program outcomes drew somewhat equivocal conclusions regarding their true efficacy because of limitations to the experimental design. However, given the paucity of

more rigorous research in the area, improved health as an outcome of increased physical activity has its own inherent return, regardless of a more favorable cost-benefit ratio defined in strict monetary terms (Bouchard et al., 1993).

## **Medical Costs and Benefits**

Estimates are made about cost to the general economy due to lifestyle-related diseases, but it is more difficult to estimate costs of an unfit work force. In total, the studies cited previously indicate that health costs would be lower for groups of employees who exercised more than other groups. More specifically, the Canadian Life study showed that medical care costs were decreased for a group who exercised compared with workers who did not (Nieman, 1995). At Prudential, disability days were reduced more than 20 percent for employees who participated in a fitness program, and the fitter workers had a 46 percent lower rate of major medical costs. Tenneco saw 48 percent lower medical costs for an exercise group compared with nonexercisers (Nieman, 1995).

A 1992 U.S. Department of Health and Human Services (U.S. Department of Health and Human Services, 1992) survey included, as benefits of a physically fit work force, improved employee morale, reduced health insurance costs, reduced absenteeism, increased output and productivity, reduced on-the-job accidents, and fewer workers' compensation claims. Other fitness benefits in the workplace are further documented by Canadian Life (Nieman, 1995), whose absenteeism rate dropped 50 percent by employees who were "high adherents" in a fitness program. At Prudential, disability days were reduced more than 20 percent for employees who participated in a fitness program. Tenneco saw a trend for fewer sick hours for exercisers versus nonexercisers (Nieman, 1995).

## **Mortality and Level of fitness**

Several independent illnesses described in this chapter have been investigated for their severity, health consequences, and possible prevention through physical activity. The overwhelming finding has been that a moderate level of regular physical activity has a very beneficial influence in reducing the risk of all these illnesses at the same time. This further begs the question then, what is the reduction in the risk of all-cause mortality relative to fitness or physical activity? A few studies have set out to answer this question.

The first investigation studied the relationship between changes in physical fitness and the risk of mortality in men. A group of researchers (Blair et al., 1995) evaluated this relationship through a prospective study of 9,777 men. The subjects were examined twice (mean of 4.9 years between initial and follow-up exams) to assess changes in physical fitness and then again assessed for mortality risk (mean 5.1 years after second exam). The main outcome measures were all-cause mortality ( $n = 223$ ) and cardiovascular disease ( $n = 87$ ) mortality. The researchers found the highest age-adjusted, all-cause death rate among the men who were physically unfit at the time of both examinations. The lowest death rate was among those who were physically fit at the time of both examinations. The men who improved from physically unfit to fit between the first and second examination had a 44 percent reduction in mortality risk relative to men who remained physically unfit between the two

exams. There was an inverse correlation between the improvements in fitness and mortality risk. Each minute increase in maximal treadmill time (level of fitness) between the first and second exams corresponded with a 7.9 percent decrease in mortality risk. The authors conclude that men who maintained or improved their fitness level were significantly less likely to die from all causes and cardiovascular disease during the follow-up than those who were persistently physically unfit.

In another investigation (Lee et al., 1999) researchers studied the independent benefits of vigorous versus nonvigorous physical activity on the risk of all-cause mortality. They investigated 17,321 men (mean age = **46** years) who were free of self-reported or physician-diagnosed cardiovascular disease, cancer, or chronic obstructive pulmonary disease. The men answered questionnaires concerning their physical activities at their baseline screening. During 22 follow-up years there were 3,728 deaths. The authors reported reductions in the risk of all-cause mortality for those who exhibited moderate levels of total energy expenditure and energy expenditure from vigorous activities. They did not find such a benefit for those engaging in nonvigorous activities. They conclude that there is a graded, inverse relationship between total physical activity and mortality. In addition, they found that vigorous, but not nonvigorous activities were associated with longevity. The researchers point out that their findings only pertain to all-cause mortality and that nonvigorous exercise has clearly been shown to benefit other aspects of health.

## Conclusion

Again, more detailed, supporting evidence of the strong association between specific morbidities (and mortality) and fitness or health follows in Appendix A. All of these investigations have generally sought to further quantify linkages between physical activity and overall health. However, as alluded to earlier, attempting to determine which types of physical activities are the most beneficial, how long and how often they really should be performed, and the specific degree to which isolated health problems may be prevented, alleviated, or at least attenuated by such activity, is very complicated because many of these variables are interrelated. Furthermore, in practice, varying levels of difficulty may be encountered when further specifying the dose-response relationship, depending on the chosen component of fitness, that is, aerobic, strength, body composition. This is primarily due to the lack of sufficient research data and/or discrete analytical methodologies, and is especially true when attempting to identify clearly defined testing cut-points on which to base standards.

---

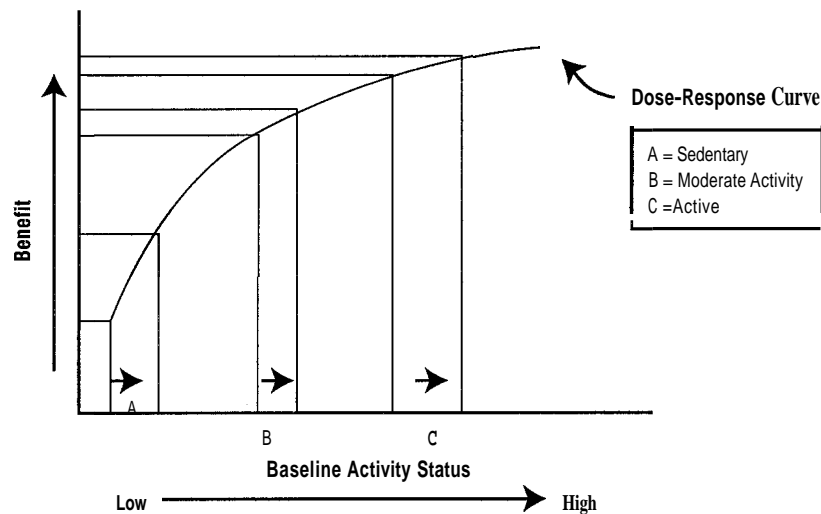
## Exercise Prescription

Since there are many lifestyle and environmental influences, as well as genetic factors that are all related to overall health (refer to Figure 2.1), it is difficult to isolate specific types and amounts of activity for each person as direct predictors of optimized wellness. However, using the available data relating to physical activity, the U.S. Government and nationally prominent health organizations have constructed general activity guidelines to reduce the prevalence and severity of disease among the population. According to a joint recommendation from the Centers for Disease Control

and Prevention (CDC) and the American College of Sports Medicine (ACSM) (Pate et al., 1995), “Every U.S. adult should accumulate 30 minutes or more of moderate-intensity physical activity on most, preferably all, days of the week” (p. 404). “One way to meet this standard is to walk two miles briskly” (p. 404). Alternatively, the ACSM has a more detailed and arguably more rigorous prescription (Pollock et al., 1998) that incorporates cardiorespiratory fitness, muscular strength and endurance, body composition, and flexibility. These prescriptions have been accepted as very general recommendations to the American public for achieving and maintaining adequate overall health, and, likely at least, modest improvement in physical capacity and performance. Other exercise prescription approaches are described later in this chapter.

When choosing an activity to perform, one must consider his or her existing fitness level. This initial baseline, along with the projected duration, are determinants of the level of exercise intensity. However the total volume (frequency  $\times$  duration) is probably most important for health benefits. Exercise adherence or consistency seems to be the dominant factor. Moreover, if the exercise is of sufficient duration, a lower or higher intensity will not alter the fact that health benefits would be gained. Although further improvements in cardiovascular endurance (physical performance) may be observed, only a minimal threshold of activity intensity is required to begin to reduce disease risk. The quality of exercise necessary for basic health is therefore based on its frequency and duration (volume) more than its intensity. Generally speaking, increased duration, frequency, along with decreased intensity, would tend to promote health over performance improvements (Pollock et al., 1998).

As described earlier, the relationship between regular physical activity and health has been studied extensively, and there is a multitude of evidence demonstrating a positive correlation (albeit nonlinear) between the two (U.S. Department of Health and Human Services, 2000). Figure 2.2 illustrates the dose-response relationship—an asymptotic continuum—between baseline activity status and derived benefit from physical activity. Clearly, the greatest health benefits are found in those people who start at the lowest fitness levels and then initiate a program of physical activity. At the highest levels of energy expenditure, there is diminishing return and potentially increased risk for injury, diminished immune response, malaise, and so forth. Further confounding the picture is the high degree of dose-response variability among individuals. From the perspective of identifying a health-related standard or specific cut-point, the general notion of a continuum is problematic. In other words, the challenge is to scientifically define a threshold of physical activity level (mode, volume, and intensity) that relates to a minimum desirable level of health and fitness. Furthermore, to develop specific fitness standards that are useful and accurate, a clear scientific rationale must be supported. In this chapter, this issue is most germane (and difficult) and will be considered in relation to a health-based standards development process.



**Figure 2.2** The dose-response curve represents that best estimate of the relationship between physical activity (dose) and health benefit (response). The lower the physical activity status the greater will be the health benefit associated with a given increase in physical activity (arrows A, B, and C). Copyright 1998 by the American Public Health Association.

## Programs and Recommendations Regarding Physical Activity

Health and fitness in the United States have received national attention since the 1970s. Many national mandates have been enacted since that decade, with increasing emphasis on how changes in lifestyle can promote health and longevity. Most of the early emphasis was on cardiovascular or aerobic exercise, but these have generally evolved into more well-rounded physical training programs. For example, the ACSM added specific strength-training guidance as late as 1990. The American Heart Association recommended endurance training alone in 1975 and added strength training in 1992. Most programs now recommend strength training in addition to endurance training, and flexibility training is also recommended in about a third of the programs. Highlighted here are excerpts from the ACSM pronouncements, and recommendations from the U.S. Department of Health and Human Services' Healthy People 2010 recommendations, the Centers for Disease Control (CDC), the American Heart Association (AHA), and the National Institutes of Health (NIH).

## American College of Sports Medicine Position Stand

In 1998, the ACSM published a Position Stand (Pollock, 1998) titled, *The Recommended Quantity and Quality of Exercise for Developing and Maintaining Cardiorespiratory and Muscular Fitness, and Flexibility in Healthy Adults*. The committee's recommendations were as follows—

#### Cardiorespiratory Fitness and Body Composition

1. Frequency *of training* —3–5 days per week
2. Intensity *of training*—55/65%–90% of maximum heart rate, or 40/50%–85% of maximum oxygen-uptake reserve or maximum heart-rate reserve
3. Duration *of training*—20–60 minutes of continuous or intermittent (minimum of 10-minute bouts accumulated throughout the day) aerobic activity. Duration depends on the intensity of the activity.
4. **Mode** *of activity* — Any activity that uses large muscles, can be maintained continuously, and is rhythmical and aerobic in nature.

#### Muscular Strength and Endurance, Body Composition, and Flexibility

1. **Resistance** training — Resistance training should be an integral part of an adult fitness program with sufficient intensity to enhance strength, muscular endurance, and maintain fat-free mass. Resistance training should be progressive, individualized, and stimulate **all** major muscle groups. One set of 8–10 exercises 2–3 days a week is recommended, with multiple sets providing greater benefit. For those aged 50 years and older, 10–15 repetitions may be more appropriate.
2. Flexibility training — Flexibility exercises should be incorporated into training to develop and maintain range of motion. Exercises should stretch major muscle groups and be performed at least 2 or 3 times a week.

The committee devised these recommendations on the basis of an extensive survey of the health-and-fitness-related literature, and compared with the earlier versions of the Position Stand, and it “has been pointed out that the quantity and quality of exercise needed to attain health-related benefits may differ from what is recommended for fitness benefits. It is now clear that lower levels of physical activity (particularly intensity) than recommended by this Position Stand may reduce the risk for certain chronic degenerative diseases and improve metabolic fitness and yet may not be of sufficient quantity or quality to improve  $\text{VO}_2\text{max}$ ” (ACSM, 1990, p. 265–266).

The ACSM views physical activity for health and fitness in the context of a continuum, acknowledging that many health benefits are accrued by going from a sedentary to a minimal level of physical activity. The authors note that, “Although the fitness paradigm that is recommended in this ACSM Position Stand is adaptable to a broad cross-section of the healthy adult population, it is clearly designed for the middle-to-higher end of the exercise/physical activity continuum.”

## Recommendations for Healthy People 2010

The U.S. Department of Health and Human Services (2000) published the goals relating to physical activity that were established for the U.S. population for the year 2010. A large committee of the country’s leading health experts established these goals in addition to many other goals relating to the quality of life for Americans. In order to focus on the key issues, they established the “Leading Health Indicators” (LHI) for each goal, including physical activity. The following is an excerpt from the LHI for physical activity—

*Regular and sustained physical activity has documented beneficial effects on cardiovascular functioning (e.g., reducing hypertension and hypercholesterolemia) but also on the prevention of osteoporosis and its sequelae (e.g., hip fractures), the effects of osteoarthritis, and on such mental conditions as depression. Physical activity is also an important element of weight control. This indicator addresses physical activity across the spectrum. . . . Although vigorous physical activity is associated with a decreased risk of cardiovascular disease, hypertension, and some cancers, a growing body of literature indicates that more moderate levels of activity can be beneficial to a person's health. Furthermore, the general population and diverse population groups are more likely to participate in moderate levels of physical activity (p. 37).*

This LHI for physical activity also calls out the following prescriptions—

*Increase the proportion of adolescents who engage in vigorous physical activity that promotes cardiorespiratory fitness 3 or more days per week for 20 or more minutes per occasion. Increase the proportion of adults who engage regularly, preferably daily, in moderate physical activity for at least 30 minutes per day (p. 37).*

These suggestions are based on the finding that only 15 percent of adults engaged in the recommended amount of physical activity in 1997, and 40 percent of adults did not engage in any leisure-time physical activity. That same year only 64 percent of adolescents engaged in the recommended amount of physical activity. The experts involved in *Healthy People 2010* declare the following health benefits from regular physical activity —

- *Increased muscle bone and strength*
- *Increased lean muscle and decreased body fat*
- *Aided in weight control and was a key part of any weight-loss effort*
- *Enhanced psychological well-being and even reduced the risk of developing depression*
- *Reduced symptoms of depression and anxiety and improved mood (p. 37)*

In addition, they report that, “The major barriers most people face when trying to increase physical activity are lack of time, access to convenient facilities, and safe environments in which to be active” (U.S. Department of Health and Human Services, 2000, p. 37).

## **re d ti from CDC /ACSM**

The Centers for Disease Control (CDC) and the American College of Sports Medicine (ACSM) collaborated on recommendations for physical activity and public health. They suggested that the current low rate of participation in regular physical activity may be due in part to the misperception that to benefit one must engage in vigorous continuous exercise. The scientific evidence clearly demonstrates that regular, moderately intense physical activity provides equal and substantial health benefits. Further, CDC and ACSM suggest that every U.S. adult should accumulate 30 minutes or more of moderately intense physical activity on most, preferably all, days of the week.



Intermittent activity also confers substantial benefits. Therefore, the recommended 30 minutes of activity can be accumulated in short bouts of activity: walking up the stairs instead of taking the elevator, walking instead of driving short distances, doing calisthenics, or pedaling a stationary cycle while watching television. The health benefits gained from increased physical activity depend on the initial physical activity level. Sedentary individuals are expected to benefit the most from increasing their activity to the recommended level.

Both CDC and ACSM also identified *two* other components of fitness here—flexibility and muscular strength—which should not be overlooked. Clinical experience and limited studies suggest that people who maintain or improve their strength and flexibility may be better able to avoid disability, especially as they advance into older age. An active lifestyle does not require a regimented, vigorous exercise program. Instead, small changes that increase daily physical activity ~~will~~ enable people to reduce their risk of chronic disease and may contribute to an enhanced quality of life. Finally, if Americans who lead sedentary lives would adopt a more active lifestyle, an enormous benefit to the public's health and to individual well-being would result.

## Recommendations from the American Heart Association

The American Heart Association (AHA) presented a statement on exercise describing physical activity and its benefits on health. They suggested that, "Persons of ~~all~~ ages should include physical activity in a comprehensive program of health promotion and disease prevention and should increase their habitual physical activity to a level appropriate to their capacities, needs, and interest (AHA, 1997, p. 4). AHA goes on to explain that—

*For health promotion, dynamic exercise ~~of~~ the large muscles ~~for~~ extended periods ~~of~~ time (30 to 60 minutes, three to six times weekly) is recommended. This may include short periods ~~of~~ moderate-intensity (60% to 75% of maximum capacity) activity (approximately 5 to 10 minutes) that total 30 minutes on most days. Resistance training using 8 to 10 different exercise sets with 10 to 15 repetitions each (arms, shoulders, chest, trunk, back, hips, and legs) performed at a moderate to high intensity (for example, 10 to 15 pounds of free weight) ~~for~~ a minimum of two days per week is recommended. (US Department ~~of~~ Health and Human Services, 1996, p. 4)*

## Recommendations from the National Institutes of Health

The National Institutes of Health (NIH) developed a consensus statement, prepared by non-advocate, non-Federal experts (NIH, 1996). It was based on the results of a conference of these experts in which investigators reported their findings regarding health, conducted question-and-answer sessions, and engaged in closed deliberations. Their consensus was as follows

*All Americans should engage in regular physical activity at a level appropriate to their capacity, needs, and interest. Children and adults alike should set a goal ~~of~~ accumulating at least 30 minutes ~~of~~ moderate-intensity physical activity on most, and preferably, all days ~~of~~ the week. Most*

*Americans have little or no physical activity in their daily lives, and accumulating evidence indicates that physical inactivity is a major risk factor for cardiovascular disease. However, moderate levels of physical activity confer significant health benefits. Even those who currently meet these daily standards may derive additional health and fitness benefits by becoming more physically active or including more vigorous activity. For those with known cardiovascular disease, cardiac rehabilitation programs that combine physical activity with reduction in other risk factors should be more widely used (NIH, 1996, p. 41 cited in U.S. Health and Human Services)*

## Comparison of the Recommendations of National Fitness and Health Groups

A more rigorous history and review of physical activity recommendations has been accomplished in the Surgeon General's report on Physical Activity and Health (U.S. Department of Health and Human Services, 1996). It emphasizes that people can increase their physical activity in many ways. It is acknowledged that both the more structured (ACSM Position Stand) and lifestyle (CDC/ACSM) approaches can work for the relatively sedentary person. Significant points of agreement across the board are as follows

- Most persons should accumulate at least 30 minutes of moderate, aerobic activity most days of the week,
- Additional health and functional benefits come with increased exercise intensity and/or volume,
- Resistance-training programs should be accomplished at least twice a week, involving 8 to 12 major muscle group exercises and repetitions each (multiple sets are not required).

However, some may still view this general guidance as too liberal or seemingly conflicting. This seems especially true with regard to the issue of intermittently (versus continuously) performed exercise.

As indicated earlier, in addition to moderately intense activity, recent consensus physical-activity guidelines include recommendations that allow for accumulating moderately intense physical activity over a 24-hour period (ACSM–CDC in the *Journal of the American Medical Association*, Pate 1995). These recommendations are based on evidence suggesting that comparable health and fitness benefits occur as long as the total amount of energy expended in moderately intense physical activity accumulated over the course of each day achieves the recommended levels, at least if the individual bouts are more than a few minutes (Pate et al., 1995). In many of the studies demonstrating a strong, inverse association between the level of physical activity or fitness and all-cause and cause-specific morbidity and mortality, the level of activity has been moderately intense and sometimes performed intermittently (Haskell, 1994; Leon et al., 1987; Paffenbarger et al., 1986; Pate et al., 1995; U.S. Department of Health and Human Services, 1996). Epidemiological data are supported by clinical studies comparing longer (traditionally 30 minutes of continuous activity) versus shorter (5 to 15 minutes) bouts of activities spread throughout the day (DeBusk et al., 1990; Jakicic et al., 1999). These studies reveal that comparable gains in cardiorespiratory fitness and various health measures occur with intermittent bouts of physical activity when the total amount of exercise is the same. Although more research remains to be done, the current consensus among major groups mak-

ing physical activity recommendations is that intermittent activity is beneficial. In addition to comparable health and fitness benefits, multiple short bouts of physical activity may increase participation and adherence (Jakicic et al., 1999). Clearly, the ACSM Position Stand is targeted at the middle to higher end of the physical activity training program continuum.

---

## The Dose-Response Relationship and Test Outcomes

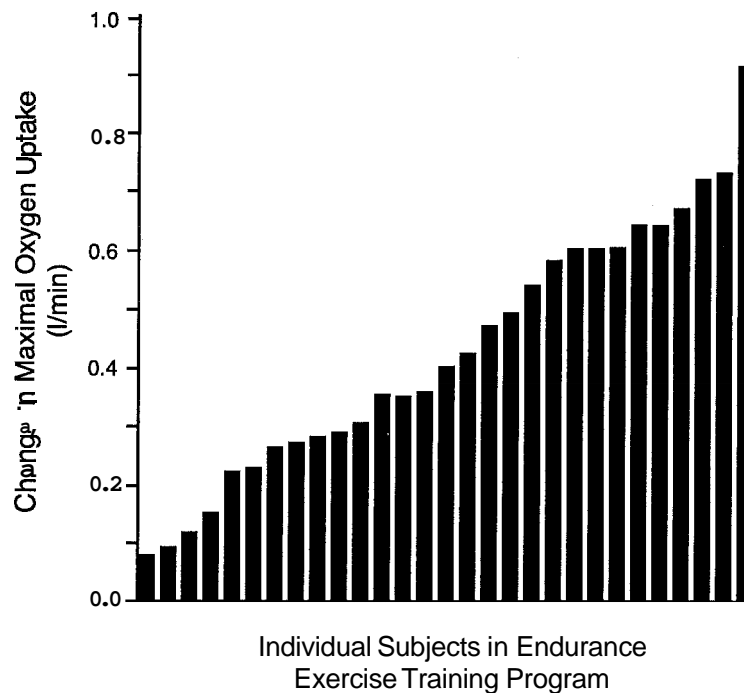
*'Everybody agrees that combat troops should be physically fit, but can we be more specific about the requirement?'* (Ramsey, *Air Force Magazine*, 1990)

The issue of dose-response might be viewed as the description of how much of what type of physical activity is required to achieve specific performance or health-related outcomes. Obviously, there is some modest disagreement among groups as to the amount of physical activity required for good health. There is consensus that regular physical activity is a part of a healthy lifestyle (which is fully compatible with Military readiness), and this habitual activity contributes to both health and fitness (Institute of Medicine, 1998). Maximum requirements of some jobs would require more rigorous physical training programs (a little more is almost always better). Remembering that the regularity of the activities, whatever they may be, is of upmost importance. On the other hand, governmental physical fitness programs have tended to focus on test outcomes (part of the response). Interestingly, it has been noted that simply enforcing past Military fitness programs has not achieved an improvement in the overall fitness levels (Institute of Medicine, 1998). It can also be argued that these testing modalities are not truly "job-related" to any reasonable extent.

Although there are many health benefits that may increase in a dose-response manner with exercise, there is presumably a point at which exercise no longer provides increasing benefits and may even be associated with increasing injury risks. The Surgeon General's Report (1996) indicated that improved health and disease prevention is associated with the recommended 30 minutes of accumulated activity on most days and increasing that activity level to some critical point can provide further benefits. Although this point is debated, it is hypothesized that the greatest benefits are seen between 30 and 60 minutes of exercise and that beyond 60 minutes the benefits plateau and the risks increase. In addition to the risk of decreased compliance due to minor injury or discomfort, there is the concern that overindulging in strenuous exercise may result in catastrophic conditions like a heart attack.

The specific dose parameters (volume, intensity, and modality) associated with health-related responses may be quite different than those typically identified when the goal is physical performance improvement (Haskell, 1994). Further, exercise prescription may be based on either a person's relative capacity or on an absolute intensity. In addition, a variety of personal characteristics should influence the dose-response relationship for any specific outcome. These characteristics would include age, gender, health status, health habits, and fitness or activity levels. Perhaps the major consideration here is the quite variable inter-individual response relative to very similar exercise prescriptions (or doses). Figure 2.3, with data from Dionne, Thibault, and Lucie (1991) shows the varied response of young men to a highly standardized aerobic training program. Figure 2.3 plots

change, after training, in maximal oxygen uptake at onset of blood lactic acid across 29 subjects. The range of improvement in maximal oxygen was greater than a tenfold difference. Finally, most of the work in the area of aerobic (versus strength) training has focused on the change in aerobic capacity, not health-related biological changes. In fact, with the exception of outcomes for reduced adiposity or insulin resistance, very little is known about the degree of health improvement (or response) to a specific dose of activity (Haskell, 1994). Further confounding the picture is the lack of linearity in responses to training and detraining as well as the larger distinctions between the acute and accumulated outcomes.



**Figure 2.3** Change in maximal oxygen uptake at onset of blood lactic acid for subjects after 12 weeks of endurance training. Reprinted, by permission, from Dionne *et al*, *Medicine and Science in Sports and Exercise*. The varied response of young men to a highly standardized aerobic training programs. 1991, 21, 177. Lippincot, Williams, & Wilkins: MD.

Another way to view this issue may be from a more qualitative approach *optimal* versus *adequate* versus *minimal* doses. An *adequate* exercise dosage, which is described as the level at or above the threshold for the desired physiological change, may be more easily identified in the literature. On the other hand, very little is currently known about the optimal or minimal dosages of exercise needed for the desired effects. Fortunately, the available knowledge is still adequate for use in exercise prescription settings directed at the general improvement of health-related outcomes (Haskell, 1994). Again, there is generally a positive dose-related versus health-outcome relationship across the entire range of exercise prescriptions, that is, more exercise is better (up to a point). Moreover, the greatest health benefits are garnered by the minimally fit after subscribing to at least a modest physical fitness program.

Finally, the amount of physical activity that corresponds to low, moderate, and high levels of cardiorespiratory fitness was explored by Stofan et al., 1998. They used questionnaires administered to a clinical population to inquire about the energy expenditure of 13,444 men and 3,972 women aged 20 to 87. The individual fitness levels were assessed by a maximal exercise treadmill test, and then compared with the reported individual energy expenditures. Average leisure time energy expenditures of 525 to 1,650 kcal/week for men, and 420 to 1,260 kcal/week for women were associated with moderate to high levels of fitness. Such energy expenditure can be achieved with a brisk walk of approximately 30 minutes duration most days of the week. The authors concluded that most people should be able to achieve these physical fitness levels. Subsequently, such activity will improve cardiorespiratory fitness enough to result in substantial health benefits.

## **Physical Activity Versus Test Outcomes/Standards**

It has been suggested that by subscribing to the ACSM guidelines, a consistently ready, fit, and healthy force would be maintained (Institute of Medicine, 1998). Although this may appear as teleologically acceptable, it is primarily a quantification of the dose and not a firm metric for a measurable outcome. Therefore, it definitely falls short as a testable fitness-related standard. This is a very important point or hurdle in the physical fitness standards development process. On the other hand, just passing a medical examination, that is, presenting with no pathologies (metabolic fitness), does not ensure compliance with a program of regular physical activity. In essence, passing a medical screen is the first (and often only) selection or retention standard to qualify for a job or career entrance.

---

## **Specific Approaches to Setting Cut-points or Standards**

Indeed the literature is very sparse with regard to methodologies and databases on which one can base cardiovascular standards. We will identify those specific efforts to define specific health-based fitness standards.

### **Aerobic**

Cooper Institute Aerobics Center Longitudinal Study: Health-Related Fitness—In this study, current U.S. Air Force minimum fitness standards (see Chapter 1 in this SOAR) were evaluated against data from a large cohort of men and women from the Cooper Institute Aerobics Center Longitudinal Study (ACLS) to determine the appropriateness of these standards as a criterion measure of health-related fitness (versus task-related fitness standards).

The ACLS cohort had accumulated about 460,000 person-years of follow-up with aerobic fitness assessment among some 41,000 men and women. The ACLS study revealed approximately 1,100 deaths. Although this particular type of analysis had not been done before, it is noteworthy that this cohort is comparable to many population-survey samples with respect to fitness levels. The

ACLS cohort served to determine, for both sexes and several age groups, standards of aerobic fitness. Anyone who attains the standard would have no more than a 50 percent greater death rate than the general population of the same age, gender, and risk profile. For example, if the mortality risk in a particular age and gender category of the general population is 2 deaths per 1,000 persons per year, the mortality risk for anyone of that same age and gender who meets the minimal fitness standard would be at most 3 deaths per 1,000 persons per year. The 50 percent cut-point is comparable to that for elevated death rates due to cigarette smoking, high blood pressure, and high serum cholesterol. These ACLS-based health-related cut-points (standards) are presented in Table 2.2. These standards are based on what is necessary to avoid a 50 percent greater mortality risk. The minimally fit cut-point (bottom quintile) used in previous studies by these researchers is a bit higher than this, but this cut-point identifies a group that is at somewhat higher risk. Due to the limited numbers of deaths occurring at younger ages, the categories presented are necessarily narrower than those for the current U.S. Air Force minimal fitness standards. Unfortunately, a similar analytical approach could not be accomplished relative to all-cause morbidity.

**Table 2.2 Aerobics Center longitudinal study health-related fitness standards**

Age (Years)	Males* (ml/kg/min)	Age (Years)	Females* (ml/kg/min)
<40	30.2		(insufficient data here)
40-49	29.3	<50	27.8
50-59	27.6	50-59	21.2

\* =  $\text{VO}_2\text{max}$

These preliminary analyses suggest that current U.S. Air Force minimum fitness standards are sufficient to promote health-related fitness, at least for males. The more limited findings for females imply that comparable U.S. Air Force standards may be too low for most women (i.e., under the age of 50). Generally speaking, persons who are physically active at the levels recommended by the recent consensus public health recommendations would be highly likely to achieve the current U.S. Air Force minimum fitness standards.

Hoeger and Hoeger (1998) have attempted to identify aerobic health-fitness standards on the basis of “epidemiological data linking minimum fitness values to disease prevention and health” (p. 7). The specific procedure used here simply identified the large decrease in mortalities observed by Blair et al. (1989) between the first (lowest) and combined second (and third) fitness quintiles as the “health threshold.” These “cut-points” related to values of maximum oxygen uptake of 35 and 32.5 ml. kg min for men and women, respectively. The authors suggest that regular exercise at approximately 50 percent of  $\text{HR}_{\text{max}}$  capacity should accomplish this training effect and qualitatively describe this level as “average.”

## Muscle Strength and Endurance

An upper BMI limit of 25 to 27 presents a fairly narrow range on which to focus as a cut-point or standard. Interestingly, the Institute of Medicine went on to recommend that a range of 25 to 30

be considered as a “caution” zone, with further Military disposition, depending on the outcome of the Service member’s physical fitness testing. The application of a health-related body composition standard here certainly has scientific support. It is also noteworthy that others (Hodgdon, 1998) have previously supported basing body composition standards on health considerations because of the lack of any scientific basis for using either performance or appearance criteria.

In-depth literature reviews and analyses have recently been completed by the Human Systems Information and Analysis Center. The topics included muscular strength and muscular endurance relative to body composition (Palmer, Rench, Carroll, & Constable, 2000). In all, no findings of published test metrics or standards were specifically linked to health-related levels of fitness. The exception to this observation might be found in Hoeger and Hoeger who at least notionally identified health fitness standards for muscular strength and endurance. However, on closer review these appear to have been chosen from a normative standpoint of around the 50<sup>th</sup> percentiles.

## Body Composition

As noted, these standards of health-related criteria may not be so simple. Perhaps the approach of establishing a health-related fitness standard lends itself better to the fitness-related parameter of body composition. A plethora of information strongly reveals the distinct relationship between BMI (body mass index) and health as a classic J-shaped curve. It is most desirable to maintain a BMI (wt/ht<sup>2</sup>) of between 19 and 25 (kg/m<sup>2</sup>), as both relative underweight and overweight are accompanied by impaired physical performance and increased risk of morbidity and mortality (Institute of Medicine, 1998). However, in the past most of the interest focused on high BMIs. It might also be argued that increasing an upper BMI limit to 26–27 would not necessarily increase all-cause mortality risks.

---

## Further Military Relevance

### Readiness and Fitness

*“The human factor in readiness and warfare always has determined the end results of hostile conflict.” (“Total Force, Total Health,” an article in Leading Edge, 1999, Air Force Materiel Command’s monthly publication)*

The number one component in readiness of our Military has been identified as qualified people (*Armed Forces Journal International*, January 1999). Herrold (Institute of Medicine, 1998) defined readiness for a Military mission as maximizing performance, minimizing unplanned losses, and adapting to changing environments. For the sake of this discussion, “readiness” will be considered the general preparedness and fitness necessary for a person to perform more than just basic Military tasks as determined by physical exercises such as push-ups, sit-ups, and running.

Interestingly, the Military Services have on occasion referred to these physical fitness measures as their physical readiness test (PRT) (Institute of Medicine, 1998). The ACSM Fitness Book (American College of Sports Medicine, 1992) describes health-related fitness as the ability to carry out daily task and unexpected bodily challenges with a minimum of fatigue and discomfort, or having the reserve to do all you want and more. Perhaps this could be described as a state of civilian readiness. Therefore, Military Services need to be a fit and “ready” force so that —

- \* Personnel will perform well under deployment or emergency conditions
- Everyday jobs can be undertaken safely and efficiently
- Costs due to absenteeism and medical problems will be minimized.

More specifically, cognitive performance as well as physical performance may be enhanced under emergency conditions if personnel are physically fit. A U.S. Army report (Pleban, Thomas, & Thompson, 1985) found that the more physically fit soldiers, as assessed on a battery consisting of chin-ups, push-ups, sit-ups, two-mile run time, and pulse rate by the Harvard Step Test, performed better on a cognitive test battery and had lower fatigue ratings during a two-and-a-half day Ranger-type sustained operations simulation.

## **Job Specific Fitness Requirements**

At first glance it might appear that the distinctions among different wartime missions specific to the various Services would enable them to be held to different fitness standards. For example, most U.S. Air Force troops would not be performing the kinds of more physical tasks that infantry battalions do. However, despite the emphasis by some on the pilot warfighter, many of the occupations within the U.S. Air Force performed on an ongoing basis as well as those that might be needed during deployment do require physical labor. Some units will be fit enough for deployment because of their everyday Military jobs such as civil engineering or maintenance units who lift and load every day. But also consider medical personnel who, when deployed, must carry heavy medical equipment as well as personal gear as they get on and off the aircraft. When on the ground, medical personnel must then be prepared to erect tents and, later, to transport patients on gurneys. Medical personnel, whose daily tasks may not be physically demanding, may thus be faced with a sudden demand to perform physical labor. They will not be prepared for physical labor by their daily work tasks. Only fitness training outside of the work area can prepare them sufficiently for deployment.

It could be further argued that the mission of any Military Service is to maintain a level of fitness year-round that will enable all personnel to perform any possible deployment task without fatigue under harsh environmental conditions when time and other stressors provide additional taxing of resources. If there is to be just one fitness standard for a Military Service member, should that standard represent the degree of fitness required to do the most difficult task to be found during deployment, under the most rigorous environmental and stress pressures imaginable? The various difficulties of instituting job-specific performance tests in the Military have been described not only in this document but elsewhere (Institute of Medicine, 1998). The attempt to institute occupation-related performance tests in the Military date back to the U.S. Army Air Corps' programs



during WWII (Institute of Medicine 1998). There are major concerns regarding this other alternative of job-specific physical performance tests/standards, for example, the potentially large number of tests the Military Services would be required to devise and administer, the frequency with which people are assigned to new occupations and/or promoted (Institute of Medicine, 1998), along with the potential lack of test sensitivity and specificity.

Moreover, performance on Military fitness tests does not correlate well with performance on task-specific job tests or even strength tests required for those Military vocations that demand at least moderately heavy lifting or carrying capacities (Institute of Medicine, 1998). A general theme throughout this SOAR is the goal of developing validated occupational performance tests, which is a complex process in and of itself. The only feasible approach to this for the Military would be to “group” jobs into testing categories. This approach, however, would tend to further confound the process in most cases.

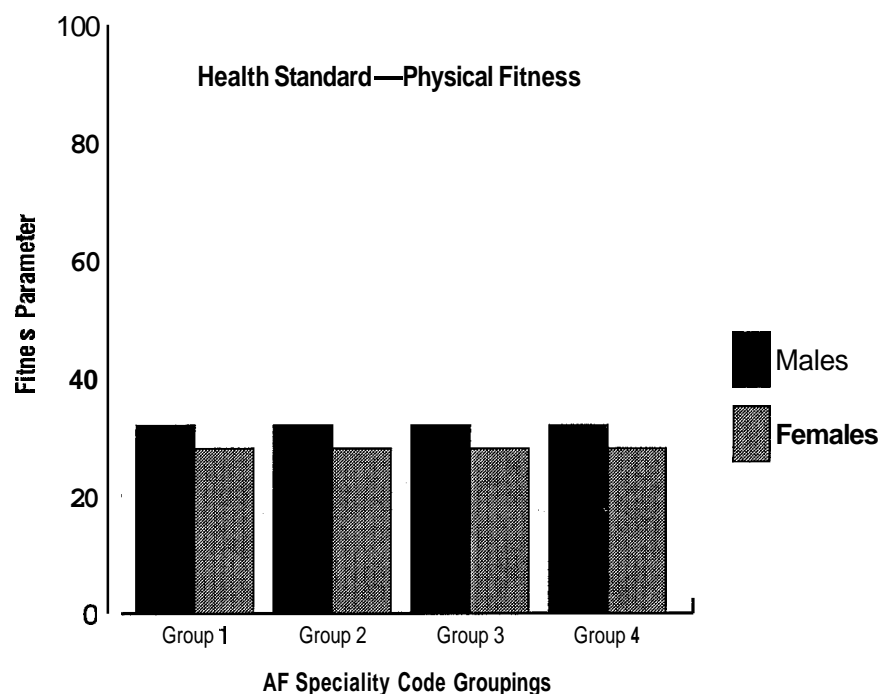
## U.S. Air Force Direction

The Military environment may present some additional challenges as well as unique opportunities. Although the Military recognizes the importance of adequate physical fitness for its members, it could be argued that a generic test to ensure full Military readiness and mission success is currently not practical or even feasible. Therefore the U.S. Air Force is exploring a two-tier approach to establishing physical fitness standards for its members. Figure 2.4 and 2.5 are notional depictions of the two types of standards. The objective of Tier I would primarily be health-based/general readiness fitness, with programs and standards that apply to all U.S. Air Force personnel. These standards would be gender-dependent to account for the physiological differences between men and women. Personnel must meet these threshold values to signify a health-related level of fitness above which distinct health benefits would be realized and identified. However, a person whose levels were below these standards would be susceptible to the increased risk of injury and disease, and decreased readiness, ability to deal with stressors, and cognitive capabilities.

Tier II of this U.S. Air Force approach would focus on an occupation-specific, performance-based fitness program that would further enhance mission readiness and accomplishment. Performance-based standards are gender independent with thresholds based on occupational requirements. These thresholds would represent each level of physical fitness necessary for personnel to meet the physical requirements of their U.S. Air Force Specialty Code (AFSC). Inability to meet these standards would imply an increased risk of mission-specific failure and greater physical fatigue. Therefore, the Tier II approach might be considered an outgrowth of the U.S. Air Force’s more limited Strength Aptitude Test (see Chapter 1), which is currently employed.

## Appearance in the Military Context

A healthy physical appearance has long been valued by the U.S. Military establishments. The 1992 book, *Body Composition and Physical Performance* (Institute of Medicine, 1992) indicates that the “appearance” rationale for body composition standards does not have a substantial relationship to performance, fitness, nutrition, or health (Institute of Medicine, 1998). However, officials from

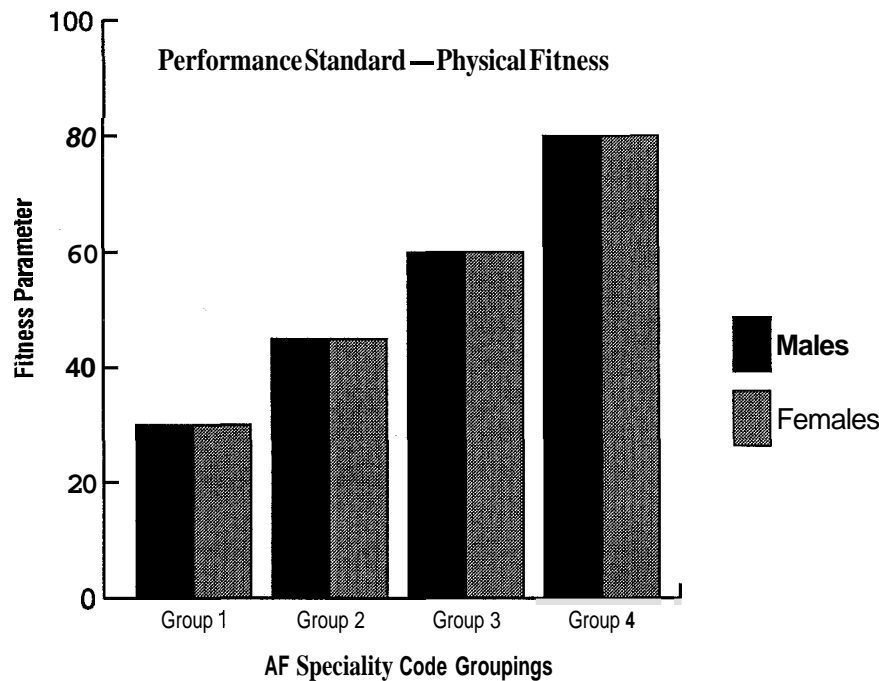


**Figure 2.4** *Notional depiction of health-based physical fitness standards (Tier I)*

two Services, the U.S. Army and the U.S. Marines, told Government Accounting Office staff that appearance was one objective of their fitness programs, stating that image is an important component of effectiveness. Since the image of a soldier is one of leanness, a fat appearance could weaken the Military image and undermine effectiveness and thus, readiness. The U.S. Navy reported that appearance is not an appropriate objective of their body fat program (Hodgdon, 1992), but rather, U.S. Navy body fat results are incorporated into a member's rating in the "Military bearing" category of officer fitness reports and enlisted personnel evaluations.

Several studies throughout the years have sought to establish the relationship between what is accepted as Military appearance and genuine measures of body composition of height and weight. A brief history of the relationship between visually judged Military appearance and actual body composition includes the work of Dupertuis (1950) who found a correlation of -0.85 between endomorphy ratings and body specific gravity. In 1952, Brozek and Keys found a mean correlation of 0.67 when subjects were rated before and after a period of semistarvation. Ward, Sutherland, and Blanchard (1976) found reliable responses of body fatness by visual appraisal, as did Blanchard, Ward, Kryzwicki, and Cannam (1979). Sterner (1984) used photographs and two raters to estimate fatness. Correlations between percent fat as documented by hydrodensitometry and that predicted from visual estimation were 0.80 and 0.79 for the two raters. Test-retest correlations were 0.93 and 0.95 for these same raters.

Subsequent to this work, Hodgdon, Fitzgerald, and Vogel (1990) conducted an experiment to determine how strongly ratings of Military appearance and fatness were associated, and to consider how reliable and valid assessments of fatness could be made in a Military population that includes personnel of both genders and various ages and races. The subjects were 1,326 U.S. Army



**Figure 2.5 Notional depiction of performance-based physical fitness standards (Tier II)**

active duty personnel, including men and women, whose body composition was established by hydrodensitometry. Appearance and fatness were rated by 11 personnel, male and female, officer and enlisted, and African-American and Caucasian. Fatness was rated on a scale from 1 (very thin) to 7 (obese) by viewing swimsuit photographs. Military appearance, using Class A uniform and swimsuit photographs, was rated on a scale of 1 (poor) to 5 (excellent). Raters were asked to use their own personal standards to assess “Military appearance.” Hodgdon et al. (1990) found that while fatness ratings could be considered valid and reliable, ratings of appearance did not fare as well. Ratings of appearance in uniform were not highly correlated with percent body fat (0.53 for males and 0.46 for females). The authors conclude that factors other than body composition, such as subjective judgment, may influence ratings and that “...it is not feasible to establish a single rating procedure which can be used to rate both Military appearance and fatness” (p. 22).

Moreover, appearance and readiness concerns may even be incompatible when maximum job performance is desired (Institute of Medicine, 1998). The U.S. Navy, for example, carried out tests that investigated the relationship between their physical readiness test and performance of materials handling tasks. Materials handling appears to be the most physically taxing task for U.S. Navy personnel. When the results on these fitness tests and two of the occupational handling tasks (box lift and box carry) were correlated with body composition measures, the correlations were generally weak. Not surprising, the one body composition measure (fat-free mass) was significantly correlated with the box lift (Hodgdon, 1990). In other words the more absolute muscle mass one has, the more weight one can generally lift. So we may have these weak relationships: body composition and visual appearance, body composition and physical performance, and appearance and performance. In fact, it has been suggested that heavier women (with high BMIs and more absolute

muscle) tend to do much better on many Military physical tasks. The conundrum then is that generally many heavier women may be better able to accomplish strenuous Military tasks but struggle to meet appearance and or body composition standards.

---

## Summary

Developing fitness standards normally involves determining some minimum level of individual physical capacities that ideally relate well to occupational job performance. For candidate selection or worker retention, the methodologies must be reasonable and legally supported as mandated by the Americans with Disabilities Act (Equal Employment Opportunity Commission, 1991). However, currently there are many workplace scenarios in which the physical requirements of the job are often quite low. Still, there may be significant interest in achieving or maintaining a reasonable level of worker fitness for a variety of reasons—general health, productivity, physical and mental readiness, and so forth. This, of course, begs the question of what specific level of fitness or physical training is desired and how it is measured. One of the scientifically supported avenues for maintaining or improving health is through physical activity or training (U.S. Health and Human Services, 1996). Thus, rather than initially attempting to require a certain level of fitness (or training) related directly to occupational performance, we have proposed an alternative or baseline, physical fitness requirement for selected applications (Military or general population): that of health-related fitness. This would be a basic level of fitness for overall health, quality of life and likely increased levels of physical performance, including both occupational and recreational activities.

The relationships between physical activity, health, and fitness are strongly related in a positive manner. These parameters tend to also reside on a continuum of effects with a distinct degree of interindividual variability. Although it is generally agreed that the degree of improvement in general health status is often closely tied to the magnitude of this improvement in fitness or physical activity, it is becoming more and more apparent that these relationships are really not simple (Haskell, 1994) but, indeed, rather complex (Bouchard & Shephard, 1993). More general, physiologic adaptations have been well documented in the literature. In fact, however, the adaptive responses to physical training (or exercise) are very complex and normally include peripheral, central, structural, and functional factors (Pollock et al., 1998). Moreover, there is really insufficient comparable data relative to the specific intensity, frequency, and volume of training to fully quantify the specific benefit outcomes (Pollock et al., 1991).

The studies on specific morbidity outcomes and physical activity or fitness are voluminous. As discussed earlier, several major reviews of these topics provide a wealth of scientific insight here, including *Physical Activity and Health—A Report of the Surgeon General* (1996), and the *Physical Activity, Fitness and Health Consensus Statement* (Bouchard, Shephard & Stephens, 1993). More specifically, other studies by leading experts have explored the effects of physical activity on a number chronic conditions and diseases. For example, these afflictions include obesity (DePietro, 1995; Stefanick, 1993), hypertension (Hagberg, 1989), cardiovascular disease (CVD) (Farrell et al., 1998; Lee, Blair, & Jackson, 1999; Stofan et al., 1998), and diabetes (Hu et al., 1999). Other studies have investigated the possible benefits of physical exercise on osteoporosis (Recker et al., 1992), cancer

(Sesso et al., 1998), clinical depression, stroke (Kiely et al., 1994), and musculoskeletal health (e.g., back pain and other injuries). Moreover, physical activity has been shown to have salutary effects on more than one condition at a time.

Besides the direct benefits of improved personal health and therefore decreased health costs, many in the corporate environment suggest that health promotion in the workplace is ultimately cost effective. Overall, a number of studies tend to suggest potentially beneficial associations between worker health, general wellness, and productivity. Although these relationships are statistically significant, the true strength of the associations may be generally low (Bouchard, Shephard, & Stevens, 1993). Therefore, some uncertainty still exists regarding the expected cost-benefit ratio when instituting an employee fitness program. However, given the paucity of more rigorous research in the area, improved health as an outcome of increased physical activity has its own inherent return regardless of a highly, more favorable cost-benefit ratio defined in strict monetary terms (Bouchard, Shephard, & Stevens, 1993).

We note that risks are also associated with engaging in regular physical activity. Although the most severe of these risks may be an increased risk of heart attack, certainly the most prevalent and critical risk of starting an exercise program would be to not continue. The scientific consensus is very clear: the greatest risk is not engaging in any physical activity at all. People at all fitness levels can safely implement the current activity recommendations for good health, tailoring a program to specific needs if necessary. If a person takes the appropriate precautions and consistently engages in physical activity throughout life, the quality-of-life outcomes can be enormous.

The recommendations for exercise prescription from a number of well-recognized sources were reviewed. A more rigorous history and review of physical activity recommendations has been accomplished in the Surgeon General's report on Physical Activity and Health. Generally, it suggests that people can increase their physical activity in many ways. It is acknowledged that both the more structured (ACSM Position Stand) or lifestyle (CDC/ACSM) approaches can work for the relatively sedentary person. Significant points of agreement across the board are that most persons should accumulate at least 30 minutes of moderate aerobic activity at least three if not most days of the week; additional health and functional benefits come with increased exercise intensity and/or volume; and resistance training programs should be accomplished at least twice a week, involving 8 to 12 major muscle group exercises and repetitions each (multiple sets are not required). It was further noted that some may still view this guidance as too liberal. This seems especially true with regard to the issue of intermittently performed exercise throughout the day.

The issue of (exercise) dose and (physical) response might be viewed as the description of how much and what type of physical activity are required to achieve specific health-related or performance outcomes. Obviously, there is some disagreement among groups as to the specific amount of physical activity required for "good health." The specific dose parameters (volume, intensity, modality) associated with improved, health-related responses may be quite different from those typically identified when the goal is physical performance improvement. In fact, with the exception of outcomes for reduced adiposity or insulin resistance, very little is known about the degree of health improvement (or response) to a specific dose of activity (Haskell, 1994).

This issue was approached from a more qualitative standpoint; optimal versus adequate versus minimal doses. An adequate exercise dosage, which is described as the level at or above the threshold for the desired physiological change, may be more easily identified in the literature. However,

very little is currently known about the optimal or minimal dosages or exercise needed for the desired effects. Fortunately, the available knowledge is still sufficient to ensure that exercise prescription settings are adequately directed at improving health-related outcomes (Haskell, 1994).

Clearly, one of the most challenging issues is correlating the exercise prescription with the expected test outcome or metric. It has been suggested that by subscribing to the ACSM guidelines, a consistently ready, fit, and healthy force would be maintained (Institute of Medicine, 1998). Although this may appear as teleologically acceptable, it generally falls short as a testable, fitness-related standard. Furthermore, it is primarily a quantification of the dose and not a metric for a measurable outcome. Similarly, just passing a medical examination, that is, presenting with no physiological maladies, does not ensure compliance with a program of regular physical activity. In essence, passing a medical screen is the first (and too often the only) selection or retention standard to qualify for a job or career entrance. On the other hand, the identification of a minimal prescription dose is an elusive and untestable.

The limited number of past attempts to “methodologically” identify specific fitness test standards or metrics were reviewed. At least for the aerobic component of fitness, the findings were really more qualitative in nature, that is, the best evidence for a cut-point would be in the low or fair percentiles from a normative population scatter. On the other hand, BMI values of 25 to 27 (wt/ht<sup>2</sup>) describe much more specifically the increased health risks associated with overweight and obesity, respectively. Little if any work has been done to elucidate health metrics for muscular strength and muscular endurance. This may well prove the most difficult fitness modality to address from a health-based perspective.

The general application of a health-based fitness standard may lend itself better to specific Military applications. A comparison of Military readiness and appearance issues with regard to health and fitness was also accomplished. The Military Services need to be a fit and ready force so that—

1. All personnel will be able to perform well under deployment or emergency conditions,
2. Everyday jobs can be undertaken safely and efficiently, and
3. Costs due to absenteeism and medical problems will be minimized.

It was concluded that a health-related fitness approach is consistent with general Military readiness considerations. Moreover, factors other than body composition, such as subjective judgment of physical appearance, are not appropriate metrics. That is, it is not scientifically feasible to establish a single rating procedure to rate both Military appearance and body fatness (Palmer et al., 2000).

We also point out that the Military environment may present some additional challenges as well as unique opportunities in the physical skills arena. Although the Military recognizes the importance of adequate physical fitness for its members, it could be argued that a generic test to ensure full Military readiness and mission success is not currently practical or even feasible. Therefore, the U.S. Air Force for one is exploring a two-tier approach to establishing physical fitness standards for its members. The objective of Tier I would be primarily health-based/general readiness fitness, with programs and standards that apply to all U.S. Air Force personnel. Tier II would focus on an occupation-specific, performance-based fitness program that will further enhance mission readiness and accomplishment. Clearly, an inability to meet these latter standards implies an increased risk of mission-specific failure, greater physical fatigue, and increased risk of injury for these career fields.

Finally, major issues have been raised in this SOAR regarding job-specific physical performance tests. Difficulties in instituting job-specific performance tests in the Military have been described not only in this report but also elsewhere (Insitute of Medicine, 1998). Examples are the potentially large number of tests the Military Services would be required to devise and administer, and the frequency with which people are assigned to new occupations and/or promoted (Institute of Medicine, 1998), along with the potential lack of test sensitivity and specificity. As a general theme throughout this SOAR, developing validated occupational performance tests may be considered a complex process in and of itself. The only feasible approach to this for the Military would be to “group” jobs into testing categories, which has its trade-offs. Further, tradeoffs between expected cost saving, available resources, and legal (ADA) issues are seen in industrial environments. Moreover, performance on Military fitness tests does not correlate well with performance on task-specific job tests or even strength tests required for those Military vocations that demand at least moderately heavy lifting or carrying capacities (Insitute of Medicine, 1998).

Organizations have sought to establish standards for job candidate assignment and worker retention to ensure job performance and safety. Incorporating standards for minimal performance on tests of physical capacity should be a scientifically defensible, sometimes lofty, goal. As an underlying theme in this text, the scientific process to establish defensible standards may be considered complex and varied. Therefore, it may not always be possible to achieve the desired degree of test validity. Rather, one must work with what is at least minimally acceptable or defensible. Furthermore, cost-benefit concerns and resource considerations may be overly constraining to the idealized outcomes. This chapter has therefore attempted to build a rationale for an ancillary approach to fitness standards development: health-based fitness levels. The scientific literature is now replete in supporting the strong association between physical activity/fitness and general heath, wellness, and quality of life. Identifying the minimal dose-response relationships, not to mention truly testable metrics or standards has proved the greatest challenge. An overview of the limited attempts thus far applied for this type of approach to the process of physical fitness standards development has been presented. However attractive or meritorious this endeavor may seem, the basic observation should be that in application or practice varying levels of difficulty may be encountered, depending on the chosen modality. This possibility is primarily due to the lack of sufficient data and/or discrete methodologies to identify clearly defined cut-points on which to base standards. Nevertheless, this should not deter further efforts to investigate or apply alternative procedures.

## Endnote

1. Readiness for all U.S. Military members is integral to the basic Military requirement of maintaining a state of readiness to go to war. In other words, it is expected that all Military members will be prepared to support the combat mission as needed, regardless of the lack of rigor associated with their primary Military occupations.

## References

1. Abbott, R. D., Rodriguez, B. L., Burchfiel, C. M., & Curb, J. (1994). Physical activity in older middle-aged men and reduced risk of stroke: The Honolulu heart program. *American Journal of Epidemiology*, 139, 881–893.
2. Air Force Fitness Program. (1997). *Air Force Instruction 40-501*. Washington, DC
3. Altrarac, J., Gardner, J. W., Popovich, R. M., Potter, R., Knapik, J. J., & Jones, B. H. (2000). Cigarette smoking and exercise-related injuries among young men and women. *American Journal of Preventative Medicine*, 18(3s), pp. 96–102.
4. American College of Sports Medicine. (1991). *ACSM guidelines for exercise testing and prescription* (4<sup>th</sup> ed.). Philadelphia: Lea & Febiger.
5. American College of Sports Medicine Position Stand. (1990). The recommended quantity and quality of exercise for developing and maintaining cardiorespiratory and muscular fitness in healthy adults. *Medicine and Science in Sports and Exercise*, 22 (6), pp. 265–274.
6. American College of Sports Medicine Position Stand. (1994). Exercise for patients with coronary artery disease. *Medicine and Science in Sports and Exercise*, 26 (3), pp. i–v.
7. American College of Sports Medicine Position Stand. (1995). Osteoporosis and exercise. *Medicine and Science in Sports and Exercise*, 27 (4), pp. i–vii.
8. American Heart Association. (1996). Fletcher, G. F., Balady, G., Blair, S., Blumenthal, J., Caspersen, C., Chaitman, B., Epstein, S., Froelicher, S., Froelicher, V., Pina, I., & Pollock, M. (2000). *Statement on exercise: Benefits and recommendation for physical activity programs for all Americans*. American Heart Association. [On-line] <http://www.americanheart.org>
9. American Heart Association. (1997). *Heart and stroke statistical update, 1997*. Dallas, TX American Heart Association.
10. American Psychological Association. (1996). *Exercise gives a lift to psychotherapy*. *Psychology in daily life, get the facts*. APA Help Center. [On-line] <http://heluinz.apa.orgdaily/erecisa.html>
11. Anderson, R., Wadden, T., Bartlett, S., Zemel, B., Verde, T., & Franckowiak, S. (1999). Effects of lifestyle activity vs. structured aerobic exercise in obese women. *Journal of the American Medical Association*, 281, 335–340.
12. Blair, S. N. (1995). Exercise prescriptions for health. *Quest*, 47, pp. 338–353.
13. Blair, S. (1999). Relationship between low cardiorespiratory fitness and mortality in normal-weight, overweight, and obese men. *Journal of the American Medical Association*, 282, 1547–1553.
14. Blair, S. N., Cooper, K. H., Gibbons, L. W., Gettman, L. R., Lewis, S., & Goodyear, N. N. (1983). Changes in coronary heart disease risk factors associated with increased treadmill time in 753 men. *American Journal of Epidemiology*, 3, 352–359.
15. Blair, S. N., Goodyear, N. N., Gibbons, L. W., & Cooper, K. H. (1984). Physical fitness and incidence of hypertension in health normotensive men and women. *Journal of the American Medical Association*, 252, pp. 487–490.
16. Blair, S. N., Kohl III, H. W., Paffenbarger, Jr., R. S., Clark, D. G., Cooper, K. H., & Gibbons, L. W. (1989). Physical fitness and all-cause mortality: A prospective study of healthy men and women. *Journal of the American Medical Association*, 262, pp. 2395–2401.



17. Blanchard, J. M., Ward, G. M., Krzywicki, H. J., & Cannam, J. E. (1979). *A visual appraisal method for estimating body composition in humans* (Institute Report No. 81). Presidio of San Francisco, C A Letterman Army Institute of Research.
18. Bouchard, C., Bray, G. A., Hubbard, V. S. (1990). Basic and clinical aspect of regional fat distribution. *American Journal of Clinical Nutrition*, 52 (5), pp. 946–950.
19. Bouchard, C., Shephard, R. J., Stephens, T. (1993). *Physical activity, fitness, and health: Consensus statement*. Champaign, IL Human Kinetics Publishers.
20. Bouchard, C., Shephard, R. J., Stephens, T. (Eds). (1994). *Physical activity, fitness, and health: International proceedings and consensus statement, proceedings of the second international conference on physical activity fitness and health*. Champaign, IL: Human Kinetics Publishers.
21. Bray, G. A. (1985). Complications of obesity. *Annals of Internal Medicine*, 103(6), pp. 1052–1062.
22. Bray, R. M., Khoult, L. A., Wheelless, S. C., Marsden, M. E., Bailey, S. L., Fairbank, J. A., Harford, T. C. (1995). *Department of Defense survey of health related behaviors among military personnel* (RTI/6019/06–FR). Department of Defense, December 1995.
23. Bronner, Y. L. (1995). Session II wrap-up: Community-based approaches and channels for controlling hypertension in blacks: Barriers and opportunities. *Journal of the National Medical Association*, 87 (6), Supplement, pp. 652–655.
24. Brozek, J., Keyes, A. (1952). Body build and body composition. *Science*, 116, pp. 140–142.
25. Burchfiel C. M., Curb, J. D., Arakaki, R., Abbott, R. D., Sharp, D. S., Rodriguez, B. L., & Yano, K. (1996). Cardiovascular risk factors and hyperinsulinemia in elderly men: The Honolulu heart program. *Annals of Epidemiology*, 6 (6), pp. 490–497
26. Campbell, W., Crim, M., Young, W., & Evens, W. (1994). Increased energy requirements and changes in body composition with resistance training in older adults. *American Journal of Clinical Nutrition*, 60, 167–75.
27. Ching, P., Willett, W. C., Rimm, E. B., Colditz, G. A., Gortmaker, S. L., & Stampfer, M. J. (1996). Activity level and risk of overweight in male health professionals. *American Journal of Public Health*, 86, 25–30.
28. Chow, R. K., Harrison, J. E., Brown, C. F., & Hajek, V. (1986). Physical fitness effect on bone mass in postmenopausal women. *Archives of Physical Medicine and Rehabilitation*, 67, 231–237.
29. Cogan, T. (1584). *The haven of health*. Reprint, London: Bonham Norton, 1596.
30. Convertino, V. A., Armstrong, L. E., Coyle, E. F., Mack, G. W., Sawka, M. N., Senay, L. C., & Sherman, W. M. American College of Sports Medicine Positions Stand (1996). Exercise and fluid replacement. *Medicine and Science in Sports and Exercise*, 28 (1), pp. i–vii.
31. Cummings, S. R. (1985). Are patients with hip fractures more osteoporotic? Review of the evidence. *American Journal of Medicine*, 78(3), pp. 487–494.
32. DeBusk, R. F., Stenestrand, U., Sheehan, M., & Haskell, W. L. (1990). Training effects of long versus short bouts of exercise in healthy subjects. *American Journal of Cardiology*, 65, pp. 1010–1013.
33. Department of Health and Human Services, National Center for Health Statistics. (1988–1994). *NHANES III third national health and nutrition examination survey*, 1988–1994.
34. Despres, J. P., Tremblay, A., Moorjani, S., Lupien, P. J., Theriault, G., Nadeau, A., & Bouchard, C. (1990). Long-term exercise training with constant energy intake: Effect on plasma lipoprotein levels. *International Journal of Obesity*, 14, pp. 85–94.

35. Dionne, F. T., Thibault, M. C., & Lucie, T. (1991). Mitochondrial DNA sequence polymorphism,  $\text{VO}_{2\text{max}}$ , and response to endurance training. *Medicine and Science in Sports and Exercise*, 21.
36. DiPietro, L. (1995). Physical activity, body weight, and adiposity: An epidemiologic perspective. *Exercise and Sport Sciences Reviews*, 23, 275–303.
37. DiPietro, L., Kohl III, H. W., Barlow, C. E., & Blair, S. N. (1998). Physical fitness, physical activity, and functional limitation in adults aged 40 and older. *Medicine and Science in Sports and Exercise*, 30, p. 55–62.
38. DiPietro, L., Williamson, D. F., Caspersen, C. J., & Eaker, E. (1993). The descriptive epidemiology of selected physical activities and body weight among adults trying to lose weight: The behavioral risk factor surveillance system survey, 1989. *International Journal of Obesity*, 17, 69–76.
39. Drinkwater, B. L. (1993). Exercise in the prevention of osteoporosis. *Osteoporosis International*, 1, S169–S171.
40. Dupreutius, C. S. (1950). Anthropometry of extreme somatotypes. *American Journal of Physical Anthropology*, 8, pp. 367–385.
41. Durbeck, D. C., Heinzelmer, F., Moxley, R. T., Schachter, J., Payne, G. H., Limoncelli, D. D., Fox, S. M., & Arnoldi, L. B. (1973). *The NASA-USPHS health evaluation and enhancement program* (Report No. N73-17091). Rockville, MD: U.S. Public Health Service.
42. Equal Employment Opportunity Commission. (1991). Equal employment opportunity for individuals with disabilities: Final rule. *Federal Register*, 56 (144), 29 CFR Parts 1602 and 1627.
43. Farrell, S. W., Kampert, J. B., Kohl, H. W., Barlow, C. E., Macera, C. A., Paffenbarger, R. S. Jr., Gibbons, L. W., & Blair, S. N. (1998). Influences of cardiorespiratory fitness levels and other predictors on cardiovascular disease mortality in men. *Medicine and Science in Sports and Exercise*, 9, 899–905.
44. Fine, J. T., Colditz, G. A., Coakley, E. H., Moseley, G., Manson, J. E., Willett, W. C., & Kawachi, I. (1999). A prospective study of weight change and health-related quality of life in women. *Journal of the American Medical Association*, 282 (22).
45. Fletcher, G. F., Balady, G., Blair, S., Blumenthal, J., Caspersen, C., Chaitman, B., Epstein, S., Froelicher, S., Froelicher, V., Pina, I., & Pollock, M. (2000). Statement on exercise: Benefits and recommendations for physical activity programs for all Americans. American Heart Association. [Online] <http://www.americanheart.org>
46. Folsom, A. R., Kushi, L. H., & Hong, C. P. (2000). Physical activity and incident diabetes mellitus in postmenopausal women. *American Journal Public Health*, 90, 134–138.
47. Gaesser, G. A. (1996). *Bigfat lies: The truth about your weight and your health*. Fawcett Columbine: New York.
48. Genaidy, A. M., Karwowski, W., Guo, L., Hidalgo, J., & Garbutt, G. (1992). A tool of increasing work tolerance limits of employees engaged in manual handling tasks. *Ergonomics*, 35 (9).
49. Gorelick, P. B., Sacco, R. L., Smith, D. B., Alberts, M., Mustone-Alexander, L., Rader, D., Ross, J., Raps, E., Ozer, M. N., Brass, L. M., Malone, M. E., Goldberg, S., Booss, J., Hagberg, J. M., Montain, S. J., Martin, W. H. III, & Ehsani, A. A. (1989). Effect of exercise training in 60- to 90-year old persons with essential hypertension. *American Journal of Cardiology*, 64, 348–53.
50. Hagberg, J. M., Montain, S. J., Martin, W. H. III, & Ehsani, A. A. (1989). Effect of exercise training in 60- to 90-year old persons with essential hypertension. *American Journal of Cardiology*, 64, 348–53.

51. Hanley, D. F., Toole, J. F., Greengold, N. L., & Rhew, D. C. (1999). Prevention of a first stroke: A review of guidelines and a multidisciplinary consensus statement from the National Stroke Association. *Journal of the American Medical Association*, 281, 1112–1120.
52. Haskell, W. L. (1994). Dose-response issues from a biological perspective. In Bouchard, C., Shephard, R. J., Stephens, T. (Eds.), *Proceedings of the second international conference on physical activity, fitness and health*. Champaign, IL: Human Kinetics Publishers.
53. Helmrich, S. P., Ragland, D. R., Leung, R. W., & Paffenbarger, R. S. (1991). Physical activity and reduced occurrence of non-insulin-dependent diabetes mellitus. *New England Journal of Medicine*, 325.
54. Hodgdon, J. A. (1990). *Body composition in military services: Standards and methods* (Technical Report No. 90–21). San Diego, CA: Naval Health Research Center.
55. Hodgdon, J. A. (1992). Body composition in military services: Standards and methods. In B.M. Marriott, and J. Grumstrup-Scott (Eds.), *Body composition and physical performance, applications for the military services* (pp. 57–70). Committee on Military Nutrition Research, Food and Nutrition board, Institute of Medicine. Washington, DC: National Academy Press.
56. Hodgdon, J. A., Fitzgerald, P. I., & Vogel, J. A. (1990). Relationships between fat and appearance ratings of U.S. soldiers (Interim Report No. 90–1). San Diego, CA: Naval Health Research Center.
57. Hoeger, W. K., Hoeger, S. A. (1998). *Lifetime: Physical fitness and wellness*. Englewood, CO: Morton Publishing Company.
58. Hu, F., Sigal, R., Rich-Edwards, J. W., Colditz, G. A., Solomon, C. G., Willett, W. C., Speizer, F. E., & Manson, J. E. (1999). Walking compared with vigorous physical activity and risk of type 2 diabetes in women. *Journal of the American Medical Association*, 282.
59. Hubert, H. B., Feinleib, M., McNamara, P. M., & Castelli, W. P. (1983). Obesity as an independent risk factor for cardiovascular disease: A 26-year follow-up of participants in the Framingham heart study. *Circulation*, 67 (5).
60. Institute of Medicine. (1992). *Body composition and physical performance*. Marriott, B. M., Grumstrup-Scott, J., (Eds.), Washington DC: National Academy Press.
61. Institute of Medicine. (1998). *Assessing readiness in military women: The relationship of body composition, nutrition, and health*. Washington, DC: National Academy Press.
62. Jakicic, J. M., Wing, R. R., Butler, B. A., & Robertson, R. J. (1995). Prescribing exercise in multiple short bouts versus one continuous bout: Effects on adherence, cardiorespiratory fitness, and weight loss in overweight women. *International Journal of Obesity Related Metabolic Disorders*, 19, pp. 893–901.
63. Jakicic, J. M., Winters, C., Lang, W., & Wing, R. R. (1999). Effects of intermittent exercise and use of home exercise equipment on adherence, weight loss, and fitness in overweight women: A randomized trial. *Journal of the American Medical Association*, 282.
64. Jones, B. H., Amoroso, P. J. (2000). Injuries in the U.S. armed forces: Surveillance, research, and prevention. *American Journal of Preventative Medicine*, 18 (3S).
65. Jones, B. H., Bovee, M. W., Knapik, J. (1992). Associations among body composition, physical fitness, and injury in men and women Army trainers. In B. M. Marriott, and J. Grumstrup-Scott (Eds.), *Body composition and physical performance, applications for the military services* (pp. 141–172). Committee on Military Nutrition Research, Food and Nutrition board, Institute of Medicine. Washington, DC: National Academy Press.

66. Kannel, W. B., Belanger, A., Dagostino, R., & Israel, I. (1986). Physical activity and physical demand on the job and risk of cardiovascular disease and death: The Framingham study. *American Heart Journal*, 222 (4), pp. 820–825.
67. Kannel, W. B., Sorlie, P. (1979). Some health benefits of physical activity. The Framingham study. *Archives of Internal Medicine*, 139 (8), pp. 857–861.
68. Kiely, D. K., Wolf, P. A., Cupples, L. A., Beiser, A. S., & Kannel, W. B. (1994). Physical activity and stroke risk The Framingham study. *American Journal of Epidemiology*, 140, 608–620.
69. Kirchmer, R. C., Lewis, R. D., & O'Connor, P. J. (1996). Effect of past gymnastics participation on adult bone mass. *Journal of Applied Physiology*, 80, 255–332.
70. Kokkinos, P. F., Narayan, P., Collieran, J., Fletcher, R. D., Lakshman, R., & Papademetriou, V. (1998). Effects of moderate intensity exercise on serum lipids in African-American men with severe systemic hypertension. *American Journal of Cardiology*, 82 (6).
71. Korkeila, M., Kaprio, J., Rissanen, A., Koskenvuo, M. (1995). Consistency and change of body mass index and weight. A study on 5967 adult Finnish twin pairs. *International Journal of Obesity Related Metabolism Disorder*, 19 (5), pp. 310–317.
72. Launer, L. J., Masaki, K., Petrovitch, H., Foley, D., Havlik, R. J. (1995). The association between midlife blood pressure levels and late-life cognitive function. The Honolulu-Asia aging study. *Journal of the American Medical Association*, 274 (23), pp. 1846–1851.
73. Lee, C. D., Blair, S. N., & Jackson, A. S. (1999). Cardiorespiratory fitness, body composition, and all-cause and cardiovascular disease mortality in men. *American Journal of Clinical Nutrition*, 69, 373–80.
74. Lee, C. D., Jackson, A. S., & Blair, S. N. (1998). U.S. weight guidelines: Is it also important to consider cardiorespiratory fitness? *International Journal of Obesity Related Metabolic Disorders*, 22 (Supplement 2), S2–S7.
75. Lee, S. W., Chaneal, P., Wheeler, G., Singh, M. (1995). *Task related physical fitness and performance standards: A Canadian forces approach*. For: NATO DRG Panel 8, Optimizing the Performance of Women in the Armed Forces of NATO.
76. Leibel, R. L., Rosenbaum, M., & Hirsch, J. (1995). Changes in energy expenditure resulting from altered body weight. *New England Journal of Medicine*, 332, 621–628.
77. Leon, A. S., Connett, J., Jacobs, Jr., D. R., Rauramaa, R. (1987). Leisure-time physical activity levels and risk of coronary heart disease and death: The multiple risk factor intervention trial. *Journal of the American Medical Association*, 258, pp. 2388–2395.
78. Malmivaara, A., Aro, T. (1995). The treatment of acute low back pain—bed rest, exercise therapy or ordinary activity? (Editorial). *Duodecim*, 111 (22), pp. 2101–2102.
79. McAuley, E., Rudolph, D. (1994). Physical activity and psychosocial outcomes. In Bouchard, C., Shephard, R. J., Stephens, T. (Eds.), *Physical activity, fitness, and health: International proceedings and consensus statement*, Proceedings of the second international conference on physical activity, fitness and health. Champaign, IL: Human Kinetics Publishers.
80. Ming, W., Kampert, J., Barlow, C., Nichaman, M., Gibbons, L., Paffenbarger, R., & Must, A., Spadano, J., Coakley, E., Field, A., Colditz, G., & Dietz, W. (1999). The disease burden associated with overweight and obesity. *Journal of the American Medical Association*, 282, 1523–29.
81. Must, A., Spadano, J., Coakley, E., Field, A., Colditz, G., & Dietz, W. (1999). The disease burden associated with overweight and obesity. *Journal of the American Medical Association*, 282, 1523–29.

82. National Institutes of Health. (1998). *Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults: The evidence report*. NIH Publication No. 98-4083.
83. National Institutes of Health Consensus Development Panel on Physical Activity and Cardiovascular Health. (1996). NIH Consensus Conference: Physical activity and cardiovascular health. *Journal of the American Medical Association*, 276.
84. National Institute of Mental Health. (1999). Depression Can Break Your Heart. Report No: 99-4592, [On-line] <http://www/nimh.nih.gov/publicat/heartbreak.cfm>
85. Nelson, M. E., Fiataron, M. A., Morganti, C. M., Trice, I., Greenberg, R. A., & Evans, W. J. (1994). Effects of high-intensity strength training on multiple risk factors for osteoporotic fractures. A randomized controlled trial. *Journal of the American Medical Association*, 272 (24), pp. 1909-1914.
86. Neiman, D. C. (1995). *Fitness and sports medicine: A health-related approach* (3<sup>rd</sup> ed.). Mountain View, CA: Mayfield Publishing.
87. Nieman, D. C. (1998). *The exercise health connection*. Champaign, IL: Human Kinetics.
88. Nieman, D. C. (1999) *Exercise testing and prescription: A health-related approach*. 4<sup>th</sup> ed. Mountain View, CA: Mayfield Publishers.
89. Nordenberg, L. (1998, July 08). Dealing with the depths of depression. *FDA Consumer*, 32 (4).
90. Paffenbarger, R. S., Hyde, R. T. (1984, Jan). Exercise in the prevention of coronary heart disease. *Preventative Medicine*, 13 (1).
91. Paffenbarger, R. S., Hyde, R. T., Wing, A. L., & Hsieh, C-C. (1986). Physical activity, all-cause mortality, and longevity of college alumni. *New England Journal of Medicine*, 282.
92. Palmer, B., Constable, S. H., & Baumgartner, N. (1997). *Expanded Air Force physical fitness battery: Muscle strength, muscle endurance, and flexibility considered* **Workshop Proceedings** (CSERIAC-PR-97-001). Prepared for the Force Enhancement and Fitness Division, Department of Aerospace Physiology and Human Performance, USAF School of Aerospace Medicine.
93. Palmer, B., Rench, M. E., Carroll, J. W., & Constable, S. H. (2000). *Health and job-specific body composition standards for the U.S. Air Force, Volume I, Final Report* (CSERIAC-00-001A). Prepared for the Force Enhancement and Fitness Division, Department of Aerospace Physiology and Human Performance, USAF School of Aerospace Medicine.
94. Palmer, B., Soest, J. (1997). *Expanded Air Force physical fitness battery: Muscle strength, muscle endurance, and flexibility considered Volume I, Final Report* (CSERIAC-RA-97-003A). Prepared for the Force Enhancement and Fitness Division, Department of Aerospace Physiology and Human Performance, USAF School of Aerospace Medicine.
95. Palmer, S. (1986). Dietary considerations for risk reduction. *Cancer*, 58 (8).
96. Pate, R. R. (1988). The evolving definition of physical fitness. *Quest*, 40.
97. Pate, R. R., Pratt, M., Blair, S. N., Haskell, W. L., Macera, C. A., Bouchard, C., Buchner, D., Ettinger, W., Heath, G. W., & King, A. C. (1995). Physical activity and public health: A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. *Journal of the American Medical Association*, 273, 402-407.
98. Pleban, R. J., Thomas, D. A., & Thompson, H. L. (1985). *Physical fitness as a moderator of cognitive work capacity and fatigue onset under sustained combat-like operations*. Report No: ARI-TR-687. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA.

99. Pollock, M. L., Gaesser, G. A., Butcher, M. D., Despres, J., Dishman, R. K., Franklin, B. A., & Ewing, G., American College of Sports Medicine Position Stand (1998). The recommended quantity and quality of exercise for developing and maintaining cardiorespiratory and muscular fitness, and flexibility in healthy adults. *Medicine and Science in Sports and Exercise*, 30 (6).
100. Ramsey, R. W. (1990). Fitness and warfighting: Combat troops should be physically fit—but can we be more specific? *Air Force Magazine*, 73 (4).
101. Ravussin, E., Valencia, M. E., Esparza, J., Bennett, P. H., & Schulz, L. O. (1994). Effects of a traditional lifestyle on obesity in Pima Indians. *Diabetes Care*, 17 (9).
102. Ready, A. E., Boreskie, S. L., Law, S. A., & Russell, R. (1993). Fitness and lifestyle parameters fail to predict back injuries in nurses. *Canadian Journal of Applied Physiology*, 18 (1).
103. Recker, R. R., Davies, K. M., Henders, S. M., Heaney, R. P., Stegman, M. R., & Kimmel, D. B. (1992). Bone gain in young adult women. *Journal of the American Medical Association*, 268.
104. Rueckert, P. A., Slane, P. R., Lillis, D. L., & Hanson, P. (1996). Hemodynamic patterns and duration of post-dynamic exercise hypotension in hypertensive humans. *Medicine and Science in Sports and Exercise*, 28 (1).
105. Sacco, R. L. (1998). Identifying patient populations at high risk for stroke. *Neurology*, 51 (3).
106. Serdula, M. K., Mokdad, A. H., Williamson, D. F., Galuska, D. A., Mendlein, J. M., & Heath, G. W. (1999). Prevalence of attempting weight loss and strategies for controlling weight. *Journal of the American Medical Association*, 282 (14).
107. Sesso, H. D., Paffenbarger, R. S. Jr., & Lee, I. M. (1998). Physical activity and breast cancer risk in College Alumni Health Study, 9.
108. Sesso, H. D., Paffenbarger, R. S., Ha, T., & Lee, I. M. (1999). Physical activity and cardiovascular disease risk in middle-aged and older women. *American Journal of Epidemiology*, 140 (4).
109. Shephard, R. J. (1992). A critical analysis of work-site fitness programs and their postulated economic benefits. *Medicine and Science in Sports and Exercise*, 24, pp. 354–370.
110. Shinton, R. & Sagar, G. (1993). Physical activity and stroke in British middle-aged men. *British Journal of Medicine*, 304.
111. Sinaki, M., Fitzpatrick, L. A., Ritchie, C. K., Montesano, A., & Wahner, H. W. (1998). Site-specificity of bone mineral density and muscle strength in women: Job-related physical activity. *American Journal of Physical Medicine and Rehabilitation*, 77 (6).
112. Sinaki, M., McPhee, M. C., Hodgson, S. F., Merritt, J. M., & Offord, K. P. (1986). Relationship between bone mineral density of spine and strength of back extensors in healthy postmenopausal women. *Mayo Clinic Proceedings*, 61.
113. Smith, E. L., Gilligan, C., Tommerup, L. J. (1994). Exercise and bone loss. *Advances in Nutritional Research*, 9.
114. Smith, R. (1995). Idiopathic juvenile osteoporosis: Experience of twenty-one patients. *British Journal of Rheumatology*, 34 (1).
115. Snow, C. M. (1996). Exercise and bone mass in young and premenopausal women. *Bone*, 18 (1).
116. Stefanick, M. L. (1993). Exercise and weight control. *Exercise and Sport Science Reviews*, 21.
117. Sterner, T. G. (1984). *Relationship between percent body fat as determined by visual estimation and underwater hydrostatic weighing*. Master's thesis, Springfield College, Springfield, OH.

118. Stofan, J. R., Dipietro, L., Davis, D., Kohl, H. W., & Blair, S. N. (1998). Physical activity patterns associated with cardiorespiratory fitness and reduced mortality: The Aerobics Center Longitudinal Study. *American Journal Public Health*, 88.
119. Thune, I., Brenn, T., Lund, E., Gaard, M. (1997). Physical activity and the risk of breast cancer. *New England Journal of Medicine*, 336 (18).
120. Tkachuk, G. A., & Martin, G. L. (1999). Exercise therapy for patients with psychiatric disorders: Research and clinical implications. *American Psychological Association*, 30 (3).
121. Tommerup, L. J., Raab, D. M., Crenshaw, T. D., & Smith, E. L. (1993). Does weight-bearing exercise affect non-weight-bearing bone? *Journal of Bone and Mineral Research*, 9 (8).
122. Total Force-Total Health. (1990). Leading Edge. *Air Force Material Command? monthly publication*, 41 (2).
123. U.S. Department of Health and Human Services. (1996). *Physical activity and health: A report of the Surgeon General*. Atlanta, GA: U.S. Department of Health and Human Services Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion. Health: International Proceedings and Consensus Statement, Proceedings of the second International Conference on Physical Activity, Fitness and Health. Champaign, IL: Human Kinetics Publishers.
124. U.S. Department of Health and Human Services. (2000). *Healthy people 2010*, 1 & 2. Washington, DC: U.S. Department of Health and Human Services.
125. U.S. Department of Health and Human Services, Public Health Service. (1992). 1992 national survey of worksite health promotion activities. *American Journal of Health Promotion*, 7.
126. U.S. Department of Health and Human Services, Public Health Service. (1996). *Physical activity and health: A report of the Surgeon General*. Atlanta, GA: U.S. Department of Health and Human Services.
127. Ward, G. M., Sutherland, T. M., Blanchard, J. M. (1976). *Evaluation of body composition of human subjects by means of visual appraisal* (Final Report Contract No. DAMD 17-74-C-4072). Colorado State University, Fort Collins, CO.
128. West, K. M. (1978). Diabetes in American Indians. *Advanced Metabolism Disorder*, 9.
129. Wier, E. K., Archer, S. L., Reeves, J. T. (Eds.). (1992). *The diagnosis and treatment of pulmonary hypertension*. Mount Kisco, NY: Futura Publishing Company.
130. Wilkinson, W. J., Kampert, J. B., Blair, S. N., Baumgartner, N., & Constable, S. H. (2000). *Physical fitness and health: A comparative review of the USAF fitness programs*. The Cooper Institute for Aerobics Research & The U.S. Air Force School of Aerospace Medicine, PE Division, USAF School of Aerospace Medicine.
131. Wilmore, J. H. (1996). Increasing physical activity: Alterations in body mass and composition. *American Journal of Clinical Nutrition*. 63.

4



# Chapter 3

## Job Analysis

**Mark Rayson, Ph.D.**

Optimal Performance Limited

Surrey, United Kingdom

---

### Abstract

This chapter describes the process of conducting a job analysis or Physical Demands Analysis, the purpose of which is to describe and quantify those aspects of physical fitness or physical performance that are relevant to job performance. Given the interdependencies between aspects of the job, the environment, and the employee, adopting a systems approach is essential in documenting and quantifying these elements. Conducting a job analysis should be an employer's first step to improve the integration of the human element into the system.

A number of techniques have been presented in this chapter to identify the most physically demanding tasks using some industrial/organizational psychology tools, and to quantify the stress and strain associated with these tasks using physiological, biomechanical, and psychophysical approaches. The strengths and limitations of the various approaches are discussed.

The issue of which approaches and techniques should be selected by the investigator will depend on many factors, including the nature of the job or task under investigation, the extent of financial and human resources available to support the work, and the expertise of the investigation team. In general terms, a multidisciplinary approach performed by a mixed gender and aged team with differing skills and perspectives is preferred, as it is more likely to elicit a complete and balanced output.

Conducting a job analysis is a complex process that requires a considerable investment of time, money, and effort. Good science and good judgment are required in equal measure. The output should provide a sound foundation for establishing occupational fitness standards, focusing physical training programs, identifying health and safety issues, and prioritizing those tasks that require job redesign. The long-term benefit to the employer of implementing these strategies will be increased productivity through improved operational effectiveness and reduced injury.

---

## Intro

### What Is a Job Analysis?

Given the focus of this State-of-the-Art Report (SOAR) on occupational fitness, this chapter concentrates on the physical demands of jobs rather than on other psychosocial aspects, normally described in other texts on job analysis. Physical Demands Analysis is a more appropriate term for job analysis in this context. Most of the published sources that discuss job analysis do so from an industrial or organizational psychology perspective, and although our focus differs, these disciplines do provide some useful measurement tools. Most ergonomics and human factors texts cover the subject, but again, a greater emphasis is usually placed on the psychosocial perspectives and less attention is afforded to the physical elements. Perhaps this is due to the lack of physically demanding occupations that remain in our increasingly automated and sedentary society (refer to Chapter 1 on the History of the Standards Development Process for more discussion on this topic). The Military, the emergency Services, and a relatively few number of jobs in the profit sector are the principal source of physically demanding occupations.

*A job analysis involves systematically collecting information about a job in order to prepare a job description. The process involves determining what tasks are included in the target job and what job skills or other employee characteristics are required. Current subject-matter experts can provide information on whether or not the tasks and skills are part of or required for the job, and their frequency or occurrence or use on the job. (Dwyer, Prien & Burke, 1987)*

Conducting a job analysis is a protracted and complex process, requiring an intriguing blend of science and judgment. Although the scientific approaches described in this chapter can take us so far, our enthusiasm for pursuing scientific rigor should be tempered by common sense and considered opinion. Achieving this balance between science and considered opinion is key to successfully performing a job analysis.

The objective of a job or Physical Demands Analysis is to describe and quantify those aspects of physical fitness or physical performance that are relevant to job performance. Given the close association and interdependencies between aspects of the job (equipment used, duration and frequency of tasks, etc.), the working environment (temperature, humidity, noise, etc), and the physical capability of the employee, a job analysis must encompass **all** parts of the system. How realistic would it be to assess the physical demands of firefighting, for example, without considering the protective clothing and breathing apparatus worn by fire fighters and the thermal load experienced during operations? It is the combined load on the human body that must be considered, not just the strength requirement to drag a hose or lift a casualty.

The description and quantification of the physical stress imposed by the job and the resultant strain imposed on the employee is a vital first step in establishing occupational fitness standards, whether they are for selection, training, or retention. The output from a job analysis provides a solid foundation from which to develop and validate job-related fitness standards. In the longer term,

understanding the physical demands of jobs and implementing appropriate fitness standards should also serve to increase productivity through improved operational effectiveness and reduced injury. Improving the match between the physical capability of the employee and the physical demands of the job is the key. The output from a job analysis may also serve to highlight training needs and to audit the effectiveness of current fitness training programs as well as to expose health and safety issues and the need for job redesign. In short, conducting a job analysis should be an employer's first step to improve the integration of the human element into the system.

There is a wide range of job analysis techniques available, some of which are appropriate to a Physical Demands Analysis. In this chapter, we attempt to encapsulate the key elements of a Physical Demands Analysis. However, our preferred multidisciplinary systems approach makes it difficult to contain within one section without omitting significant topics. We have attempted to find a reasonable compromise. Following an introduction, an overview of some of the more useful techniques is provided, first, for identifying the most physically demanding tasks using some industrial/organizational psychology tools, and second, for quantifying the stress and strain associated with these tasks using various techniques that we have classified under the titles of physiological, biomechanical, and psychophysical approaches.

## Who Should Conduct a Job Analysis?

A job analysis is often best conducted by an interdisciplinary team comprising in-house workers and supervisors who are highly familiar with the jobs under scrutiny, and external consultants who can bring greater objectivity and reliability to the process. It is useful to have a multiethnic and multigender team because different perspectives on job and task performance are often valuable.

Historically, defining the requirements of the most physically demanding occupations has involved investigating the performances of men, since it is men who hold, or at least held, most of these positions. However, this approach may be problematic because men and women perform tasks differently (Courville, Vezina & Messing, 1991; Stevenson et al., 1990). Further, occupations and equipment have usually been designed by men for men, thus creating a systemic bias against women in equipment, payloads, and work organization (Courville et al., 1991). These potentially biasing factors should be borne in mind when conducting a job analysis, and efforts should be made to counter them.

## Which Employees Should be Evaluated?

*"To calculate sample size for a Physical Demands Analysis we need to know the statistical properties of the process generating the data, the width of the confidence interval, and the probability of the interval containing the true value" (Wilson & Corlett, 1995).*

In reality, this information is usually unknown, and selecting a suitable sample of employees to investigate will always be a compromise between recruiting the numbers and variety of personnel to fulfill criteria for a valid statistical analysis and recruiting personnel who are available and will-

ing to take part. The extent of resources, both human and financial, will also inevitably and unfortunately impinge upon the sample size.

In simple terms, the greater the sample size is, the greater the accuracy of the data will be. Although there is one argument to select “normal” employees and “normal” work scenarios, there is another to opt for extremes of the employee population. If we select “normals” we may miss completely some of the minority groups in the work force and some of the more unusual, less frequent tasks that might contain greater physical demands. The preferred approach will certainly involve adequate representation from key minority groups (usually women and ethnic minorities), though absolute target numbers will depend on the approach adopted and the resources available. Further details about sampling (such as whether it should be random, stratified, or clustered) may be found in more specialized texts (e.g., Sinclair 1975, Ferguson & Takane, 1989).

Data on employees should be recorded anonymously to prevent the matching of particular data to particular people. This can usually be achieved fairly easily with paper records by allocating employees a unique identifying number and using the number only on paper and computer records. However, it may not be as straightforward with video and photographic records. As a minimum, all participants should be briefed on the objectives of the investigation and how their data will be protected.

---

## Identifying the Physically Demanding Tasks and Elements

There are two broad approaches to conducting a Physical Demands Analysis. The first is to focus on the work itself, describing the purpose of the job and the equipment used, and the second is to focus on the employee, describing the physical and behavioral requirements of the job. Similarly, there are many techniques that can be employed to elicit information, organize it, and deploy it to make decisions about the physical job demands. These include observation, interview, questionnaire as well as investigating other written material that may be available.

In reality, it is good practice to use a combination of techniques, since this should provide a more holistic and valid outcome. The selection of techniques will depend on a number of factors, including the nature of the job under investigation, the employee’s ability and willingness to understand and tolerate the different techniques, the experience and the preferences of the user, the physical environment in which the analysis must be conducted, and, inevitably, the resources available. Some techniques require verbal or written input from employees, others are observation based and require minimum input from, or interruption to, the work force.

## General Information Gathering

Before implementing these measurement techniques, it is advisable to carry out an initial familiarization and documentation phase in which the person conducting the Physical Demands Analysis gets a grip on the job and gathers relevant documentation, since this will steer subsequent stages of the process. The objective of this preliminary phase is to establish what the job entails, its objectives, and its task elements. A relatively small effort on this phase may avoid much subsequent nugatory activity.

Materials such as a job description, written procedures, training manuals, work rosters, shift schedules, and relevant reports covering any problems, accidents, or legal cases that concern the job or task under scrutiny should be collated. These and other sources of relevant information may be available from the Human Resource or Occupational Health Services if they exist. In particular, any records on sickness absence and musculoskeletal injury associated with employees in the job categories of interest may provide useful insights into problem areas.

Although these sources of information are too superficial to use alone, they can guide more detailed investigation. They also help to ensure that all of the critical aspects of the job (i.e., a proper representation of the job) will be covered. Employees at different levels within the organization, including job incumbents, supervisors, union representatives, and managers, should conduct checks for completeness and relevance of all work tasks.

## Task, Environmental, and Human Factors

An employee's ability to fulfill the requirements of a job depends upon three inter-related components—the task itself, the environment in which the task is performed and the capability of the worker. Each of these three components has a number of elements. It is these elements that must be documented and quantified in a job analysis using the various techniques and approaches that are described in subsequent sections. Describing these components and elements in detail is beyond the scope of this chapter, but an outline of these topics is provided below. For a more detailed discussion the reader is referred to the textbook and article by Ayoub & Mital (1989) and McDaniel (1998), respectively.

**Task elements**—The task elements describe the mode, frequency, intensity, and duration of the task, the postures, any objects involved in the task and any equipment that is used.

The mode or type of activity is important as some activities are more readily sustainable than others, or involve a greater or less stress or strain. The main source of variation is probably the amount of muscle mass involved in different activities and whether the effort involves static or dynamic muscle activity. For example, higher levels of work are sustainable while cycling compared with lifting (Petrofsky & Lind, 1978a) and while running compared with loaded marching (Rayson, Bell, Davies, & Rhodes-James, 1995). The greater the frequency and intensity of a task, the greater is the stress of that task and the greater is the resultant strain on the employee. Thus documenting the mode, frequency, duration, intensity, and rest periods is a critical aspect of quantifying the demands of the job.

Posture refers to the position the body adopts to initiate an activity. The same activity can often be performed in different postures, e.g., lifting in a stoop (straight legs), squat (straight back) or free-style (semi-squat). Squat posture is biomechanically least stressful in lifting tasks, but the stoop posture leads to lower energy expenditure. The free-style posture is considered least stressful or least tiring.

A variety of factors associated with any object that is handled can affect task performance and hence these factors should be documented. These include the dimensions, symmetry and the presence or absence of handles (Ayoub and Mital, 1989). The dimensions of objects have an effect on handling

capacity, energy expenditure and spinal stresses. Handling loads with handles is safer and less stressful and may allow employees to handle approximately 15% more load (Snook & Ciriello, 1991).

The ability to handle load is also a function of the vertical height and distance of lift (Snook, 1978a and b). Lifting capability decreases with vertical height above the ground. The decrease in capacity may be as large as 23% when the starting point of the lift is changed from knuckle to shoulder height for example. The height of lift also affects the stress on the spine. Lifting from the ground is more stressful than lifting from knee or hip height. Asymmetrical loads are more stressful (intra disc pressures and shear forces increase) to lift therefore reducing capability (Mital and Fard, 1986).

**Environmental elements**—The environmental elements describe the ambient temperature, humidity, altitude, noise, air pollution, work space, clothing etc. Core temperature is maintained in humans over a relatively small range between approximately 36 to 40°C. Outside of this range, thermal regulation is impaired or even lost. During activity, heat production is greatly increased, and if this heat is not dissipated, core temperature increases rapidly to intolerable levels. During the wearing of protective clothing or in hot, humid climates, heat loss mechanisms are impaired (Kolka, 1992).

Even under a relatively modest heat load, heart rate, core temperature and sweat rates increase, and work output declines. For example, in one study where the temperature was increased from 17 to 27°C, lifting capacity declined by 20%, pushing capability by 16% and carrying capability by 11% (Snook & Ciriello, 1974).

Restrictions in working space on posture, movement and working capacity have been documented both in terms of increased physiological and biomechanical strain and perceived discomfort ratings (Mital, 1986). Subjects experienced a 13% decline in carrying capabilities when loads had to be carried through a 56cm wide passage, for example. Often both the body and load have to be re-oriented resulting in slower and cautious movement. Limited headroom also reduces lifting capacity (Ridd, 1985).

**Human elements**—The human elements describe the somatic factors, such as gender, age, body dimensions; psychic factors, such as attitude and motivation; sensory factors such as perception, integration, and transmission of information; health and physical training state. Of particular interest to us due to our concern with occupational fitness are the somatic and health and fitness factors. These factors although not strictly relevant to a job analysis (they do not define the job itself) are of relevance to a Physical Demands Analysis as they impact on the employee's ability to perform the job.

Measures of body size and composition are well documented as predictors of Military task performance, including lifting capability (Nottrodt & Celentano, 1987; Rayson, Holliman & Belyavin, 2000), carrying capability (Rice & Sharp, 1994) and loaded marching (Frykman & Harman, 1985; Rayson, Holliman, & Belyavin, 2000). In general terms, larger employees with greater muscle mass and less body fat perform physical demanding tasks more effectively, though the exact relationship between these measures and performance varies according to the details of the task.

Fitness scores, especially on muscle strength, muscle endurance, muscle power and aerobic capacity have long been shown to be positively associated with job performance (e.g., Sharp et al., 1980; Rayson, Holliman, & Belyavin, 2000) and negatively associated with risk of injury (Chaffin, 1974; Jones, Bovee, & Knapik, 1992; Harwood, Rayson, & Nevill, 1999). Indeed, it is intuitive to

expect that strong and aerobically fit employees are more likely to perform the job effectively, have a greater reserve capacity and hence are less susceptible to injury.

The influence of age and gender on job performance is relevant insofar as women and older employees typically have a lower physical capability and therefore a lower work tolerance and less reserve capacity and their male and younger contemporaries. However, recent implementation of so-called gender-free job-related physical selection standards by the Armed Forces in some countries such as the United Kingdom (Rayson, Pynn, Rothwell, & Nevill, 2000), has reduced the relevance of age and gender of employees. In theory, job-related selection standards will ensure that all employees have the required physical capability to meet the requirements of the job and hence age and gender afford less significance.

## Observation

Careful observation of work performance can provide data on the occurrence, frequency, and duration of specific activities. This technique is most useful in a Physical Demands Analysis, as opposed to other aspects of job demands (e.g., cognitive and social) as it depends on visual activities, and the physical components are normally readily visible. Some, such as static efforts may not be, however.

*All observational methods have the deceptive appearance of simplicity, giving the potential user the impression that their use is easy and their results simple to determine and conclusive. Unfortunately, this is not so, and potential users should be aware of the need for training in the method, monitoring of its use and supporting knowledge for the effective application of its results. (Wilson & Corlett, 1995)*

Observation is ideally suited to jobs or tasks that involve short and repetitive work cycles. In these types of activities, several hours of observation may suffice to capture the entire extent of the physical requirements. By contrast, jobs that are not structured or repetitive may require extended periods of observation before a single relevant incident is observed. Similarly, tasks with variable physical demands may not be best suited to this technique. A law enforcement officer making an arrest provides an example of such a task. It may require several days of observation before a single incident is captured, and the variability surrounding the manner in which this task is carried out, may require many repeated exposures before “typical” modus operandi can be assured.

If the task is conducted quickly, or involves complex or highly skilled movements, it may not be possible for the observer to keep up with events, and direct observation may not be appropriate. The use of video recording and subsequent data analysis may overcome this problem. Videotaping may also overcome another potential weakness of this technique—the presence of an observer interfering with the normal pattern of behavior. Often, observation is best deployed alongside other information-gathering techniques such as interview or questionnaire.

Activity sampling is a time-structured observational approach in which snapshots of activity are coded at predetermined time intervals, which might range from 10 seconds to 1 hour. The objective of activity sampling is to quantify the proportion of time spent performing different activities often with a view to focusing further investigation into the more frequent tasks. In preparing to use

this approach, there are four issues that need to be considered, which include the classification of activities, the development of a sampling schedule, information collection and recording, and the actual analysis of activity samples. These are described in detail by Kirwan and Ainsworth (1992) on pages 42 to 44, but a brief description of each of the four issues follows.

To ensure that all of the useful activities are recorded, observers need to familiarize themselves with the activity and ideally pilot the data collection procedure, before data collection commences in earnest. Up to 20 discrete activities can be coded unless video is used, in which case any number can be managed. The activities targeted must be clearly distinguishable and the limits identified (i.e., the start and end).

The sampling interval can be calculated according to the Nyquist criterion—the interval between samples must be less than or equal to half the duration of the shortest activity that is being coded. For example, if a gunner in a tank must load and fire a shell every 1 minute, replenish shell supplies every 2 hours, and change/fill the fuel tanks every 8 hours, a sampling interval of 30 seconds or less would be required over a number of days. Either a fixed or a random sampling interval can be adopted. Fixed is normally preferred unless the tasks are of very long duration, or unless the repetition rate is very high, in which case fixed sampling can lead to a systematic bias. Sampling is normally continued for sufficient duration to sample the full range of activities and should yield approximately 1,000 sampling points for each session.

Before the measurement phase commences, the employee should be briefed and asked to perform the work as normal. The level of detail recorded can vary from a simple tally in which the occurrences of a particular activity are recorded (but only frequency will be elicited) to recording the activities sequentially, in which case both frequency and duration will be obtained.

Kirwan and Ainsworth emphasize the need for observers to devote sufficient time to ensure they are familiar with the task under investigation and to check that the task categories are clear, exhaustive, and mutually exclusive. Carrying out a small pilot study will highlight any unforeseen problems. A category for “other activities” should provide a useful catch-all, at least in the pilot study. More categories may need to be defined if uncoded activities keep recurring. A programmable beeper should be used to prompt data collection. If video is used, a time stamp can be recorded on the image.

Recently, we successfully used handheld computers in the field to log information as it occurs (unpublished data). Activity, posture, terrain, and events were logged directly onto the computer using drop-down menus arranged in a hierarchical manner. Additional information can be recorded in free text either directly onto the scribble pad on the small screen or orally on minitape or disk recorders. At the end of the recording session the data on the palmtops are downloaded into an Access database for later merging and analysis.

Observation is a useful means of identifying the sequence of activities involved in a task as well as documenting in an objective manner the frequency and duration of activities. We have found it useful to verify the accuracy of other forms of data—sometimes the way in which the job is actually performed differs significantly from the description in the training manual or user instructions. Observation can also uncover additional activities that had not been foreseen. Care must be taken when analyzing the data to distinguish between frequency and criticality. Just because an activity is frequent does not necessarily mean it is an important aspect of job performance. Similarly, a critical task may only be performed once a month, but if the employee’s life or a colleague’s life depends on it, the fact that the task is performed infrequently may be irrelevant.



## Questionnaire

The use of questionnaires can be a cost-effective technique to elicit information, though self-completion questionnaires are only appropriate for personnel who can verbalize well on paper. Questionnaires can be distributed personally or by post and can be completed at any suitable time and location. They can be structured with precoded responses, or be “open,” allowing an employee to write down a response in his or her own words. There are advantages and disadvantages with both variants, but one major advantage of the precoded version is the ease of data collation and analysis, as well as a lower employee burden. Questionnaires can also, of course, be administered by an interviewer, overcoming some of the limitations with the self-completion approach.

Most of the questions in questionnaires hinge around two major issues: the relative importance of the task to the job and how often the task occurs (criticality and frequency). Some of the better known questionnaires used for job analysis, such as the Position Analysis Questionnaire (PAQ) (McCormick, Jeanneret, & Mecham, 1972), the Occupation Analysis Inventory (Cunningham, Boese, Neeb, & Pass, 1983), and the Work Profiling System (Saville & Holdsworth, 1989) are large, time-consuming questionnaires that are not well suited to investigating physical behaviors. They take too long to administer and elicit large amounts of information that are, by and large, irrelevant to our needs.

However some job-orientated techniques, such as Fine and Wiley's (1971) Functional Job Analysis (FJA) (see also Fine & Cronshaw, 1999) and Annett, Duncan, Stammers & Gray's (1971) Hierarchical Task Analysis (HTA) can be useful. FJA is a technique used to describe what employees do in standardized language. The focus is on the tasks they perform (i.e., purposeful actions organized over time to address an objective). HTA analyzes the tasks, dividing them into increasingly specific subtasks in a hierarchical fashion. An alternative employee-orientated technique is Repertory Grid Technique (Kelly, 1955). Its advantage over some of the other employee-orientated approaches (e.g., PAQ) is that the employee is not limited in his or her response by prestructured categories.

Other techniques such as the Task Ability Scales (Fleishman & Quintance, 1984), the Threshold Traits Analysis (Lopez, 1986), and the Minnesota Job Requirements Questionnaire (MJRQ) (Desmond & Weiss, 1973 & 1975) can be useful during a Physical Demands Analysis. Fleishman's Task Ability Scales include psychomotor characteristics. Their relevance to successful performance on specified tasks is rated using graphic 7-point scales. Three points on the scale are anchored by examples of concrete action. Lopez's tool measures the relative importance of 33 characteristics spanning 5 main attributes, including physical characteristics. Brief work-oriented definitions are provided for each characteristic. The interviewee is asked the importance of the characteristic as a job requirement and the weight it has for total work performance. The MJRQ is a short questionnaire with 45 items, some of which encompass physical ability (e.g., “precise movement of fingers in the handling of very small objects”). The employee has to indicate on a 7-point scale the importance of the specified actions or activities. For further details consult the original papers or for an overview refer to Drenth, Thierry, & de Wolff (1998).

The checklist is an example of a precoded questionnaire in which the employee normally indicates by circling or scoring whether a particular event does or does not occur. However, with such closed

measurement vehicles, adequate attention must be devoted beforehand to ensure that the correct issues are being addressed, the right questions are being asked, and the right language is being used.

An example of this type of approach applied to a Physical Demands Analysis would be to design and implement a task-specific inventory that includes work-task statements that clearly define specific work tasks. The task list might be drawn up by the investigator and finalized by observation and discussion with employees and supervisors. An example of task descriptions used by Prof. Tony Jackson in an oil production facility is provided below. The task list is then administered to a random sample of workers who rate each task in terms of its importance (criticality), frequency, time spent, and sometimes difficulty. Difficulty can be assessed using the rating of perceived exertion scale (Borg, 1985) or the physical effort scale (Fleishman, Gebhardt, & Hogan, 1984). A weighted index can then be calculated for each task on the basis of the relative priorities attributed to importance, frequency, and time spent on those tasks that are classified as the most difficult. Those job tasks that are identified as key tasks then form the basis of any further initiatives, such as the development of selection or retention tests.

An alternative to inventories and checklists is activity diaries. Asking employees to maintain activity diaries that document their activities during predefined periods (e.g., typically in 30- or 60-minute blocks) can provide useful information. However, the burden on the employee is quite high, and the accuracy and usefulness of the information is variable.

Questionnaires are also available to gather data on the incidence, prevalence, and causes of musculoskeletal injury in the work force. The Nordic Questionnaire (NMQ) (Kuorinka et al., 1987) provides one suitable example. After a personal details section, general survey questions provide indications about prevalence and disability. Different sections covering four separate body areas follow to establish the severity of any disorder. A general health section rounds off the questionnaire.

### **Worked Example 1**

Task list developed for an oil production facility

1. Breaking—turning valves at wellhead
2. Breaking—turning valves in awkward positions
3. Breaking—turning 4" to 6" valves (low pressure)
4. Breaking—turning 2" to 4" valves (high pressure)
5. Lifting—carrying valves and flanges
6. Lifting—carrying pumps and motors
7. Manipulating heavy fittings in awkward positions
8. Lifting 30–50 lbs. to waist height
9. Lifting 50–75 lbs. to waist height
10. Lifting 75 to 100 lbs. to waist height
11. Lifting 30–50 lbs. to above waist height
12. Lifting 50–75 lbs. to above waist height

## Interview

An interview is normally conducted with one person but can be conducted with a group. A series of questions are asked, and the responses are recorded either on paper or directly punched into the computer. Alternatively, the whole process is recorded on tape or video for later analysis. Interviews can be either structured, whereby the questions are specified in advance (e.g., orally administered questionnaires), or unstructured, whereby only a general topic is defined and the interviewer encourages the respondent to talk about anything within that topic.

The Critical Incidents Technique (Flanagan, 1954) involves asking job incumbents or supervisors to provide examples of effective and ineffective performance. The idea is that people remember critical incidents even though they occur relatively infrequently. The focus is on both positive and negative events that have a potentially important effect on system objectives. Once several hundred examples have been collected, they are categorized into 10–20 groups, via several reiterative processes by knowledgeable employees (knowledgeable about the job concerned). The resulting categories, in our case of physical dimensions, are deemed to be relevant to effective job performance.

This open-ended technique is best employed early in the job analysis as it sometimes uncovers key problem areas in a cost-effective manner. It can simply be used in an open question such as “describe an incident which occurred to you or another employee you were watching while launching the rescue boat/constructing a bridge/hauling up a ladder.” Or it can be used in a more systematic manner by interviewing employees on a regular (weekly, monthly) basis or by asking employees to respond anonymously by filling in and returning forms. The advantages of this technique include its ability to pick up rare events that might not be uncovered by other techniques (e.g., observation). The disadvantages lie in its reliance on accurate memory and reporting by people.

The previously mentioned Repertory Grid is a similar interview method in that it too tries to elicit specific examples of successful and unsuccessful performance. However, the technique focuses on the person rather than the task by asking the employee to identify ways in which an effective employee differs from an ineffective employee. The interviewer explores examples of task performance and their relationship with the types of physical attributes. For example, which physical characteristics does the effective employee have that the ineffective employee lacks?

---

## Approaches to Quantifying the Physical Demands of Jobs

This section outlines different approaches for quantitatively assessing the physical stress (the demands) and the strain (the person’s response) associated with work tasks. Obtaining data on factors such as posture, force, and intensity of work assists in the understanding of work demands by quantifying them. These data may be used for assembling a battery of candidate fitness tests, for identifying tasks or jobs in need of redesign, for comparing the demands of jobs before and after job redesign, and for setting objective targets for rehabilitating injured personnel.

Although simple generic task work rates can be derived from tables or predicted using equations, more complex occupational tasks must be uniquely measured. There are three basic meas-

urement approaches to quantifying job demands — physiological, biomechanical, and psychophysical (Ayoub & Mital, 1989). Again, including more than one approach is preferable to avoid bias in perceptions and drawing incorrect conclusions about the demands of an occupation.

## Physiological Approach

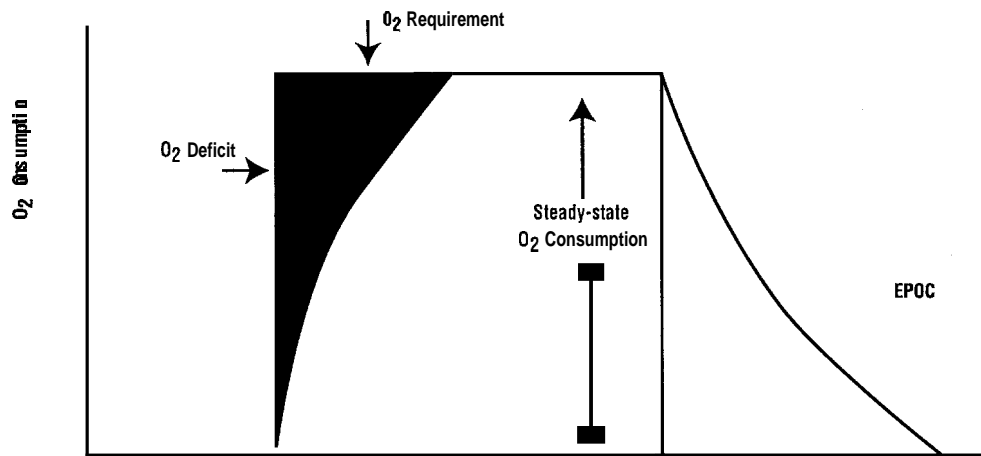
The physiological approach appraises the strain on the cardiovascular and respiratory systems through the measurement of responses such as heart rate (HR), oxygen uptake rate ( $\text{VO}_2$ ) or lactate accumulation. During dynamic activities, such as walking or running in which the primary energy supply is via aerobic metabolism, HR and  $\text{VO}_2$  are linearly related to the work performed, so the intensity of the work can be estimated by measuring either HR or  $\text{VO}_2$ . During static activities, such as maintaining a posture or holding an object, the physiological responses may be different from dynamic activities since the HR response to static exercise is largely independent of the bulk of muscle involved.

Semidynamic activities, such as lifting and carrying, include both dynamic and static components. The static component in these tasks may be significant, both in grasping the object to be lifted and in postural control. These static components may be integral to performing the task but they may not contribute to accomplishing the work (i.e., raising the mass a given height). Thus, the physiological responses to the static components are superimposed on the responses to the dynamic components.

Anaerobic metabolism and the role that anaerobic metabolism plays in meeting the energy demands of work tasks is a complex subject. At a simplistic level, anaerobic metabolism plays a significant role at the onset of dynamic activity before the cardiovascular and aerobic energy systems have time to catch up with the work demand, and an oxygen deficit is incurred, and during intense dynamic activity in which the energy demands outstrip the ability of the aerobic system to meet the requirement (Figure 3.1). Anaerobic metabolism can result in the accumulation of lactate and mark the onset of fatigue and ultimately the cessation of activity. Lactate accumulation occurs when the rate of production exceeds the rate of removal, either due to the sheer volume being produced or to the reduced blood flow and subsequent impaired removal. Lactate accumulation results in perceptions of fatigue and reduced contractile ability of the muscle, forcing the employee to stop or slow down until pH rises again and a reasonable level of homeostasis is restored.

The contribution of anaerobic metabolism to the overall metabolic demand of the activity can be estimated by measuring the  $\text{VO}_2$  (see Oxygen Uptake, below) during task performance and during the recovery period, for say 15 minutes. From a plot of the time versus  $\text{VO}_2$  as shown in Figure 3.1, the area under the curve can be calculated by numerical integration, and resting  $\text{VO}_2$  subtracted. The relative contribution of aerobic and anaerobic metabolism can then be calculated by comparing the areas under the curve during the work and recovery periods.

This cycle of exercise, fatigue, and recovery depends on the intensity and duration of the activity and recovery periods and the fitness of the employee, so it is vital that **all** of these aspects of performance are appropriately controlled and monitored. All too often published papers on the physical demands of specific occupations fail to document how participants were selected, how representative participants were of the general work force, and how the rate of work was established and controlled.



**Figure 3.1** Illustration of the oxygen deficit at the onset of activity and the excess post-activity oxygen consumption (EPOC), indicating the interaction of aerobic and anaerobic energy pathways required to meet energy demands. Reprinted by permission from J. H. Wilmore & D. L. Costill, 1999, *Physiology Sport and Exercise*. 2<sup>nd</sup> ed. (Champaign, IL: Human Kinetics), 135.

**Oxygen Uptake** — Energy expenditure during dynamic work is usually expressed as a rate (i.e.,  $\text{kJ}\cdot\text{s}^{-1}$  of  $\text{VO}_2$ ; where 20.6 kilo Joules equals 1 liter of oxygen). Depending on the mode of activity, it is sometimes more appropriate to adjust the rate of work by body mass (e.g.,  $\text{VO}_2$  in  $\text{ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ ). Direct assessment of the amount of energy produced by the body is not normally possible, so the indirect assessment of energy cost via the measurement of  $\text{VO}_2$  is often used as the next best alternative.  $\text{VO}_2$  during aerobic exercise is directly proportional to energy expenditure. The mechanical efficiency (the external work produced divided by the total energy produced) varies somewhat between persons, so the more complex and skilled the task is, the greater the variation will be. For common tasks such as walking and running there is less interindividual variation, and therefore the measurement of small numbers of employees normally suffices. For the more skilled and complex tasks, greater numbers may be required. Where there is a significant thermal load, static components to the task, or anaerobic contribution, additional measurements are required to fully encapsulate the physiological demands (e.g., HR, lactate, or body temperature).

Obtaining  $\text{VO}_2$  data for specific jobs and tasks enables the most aerobically demanding tasks to be identified, quantifies that demand, and enables any job modification to be evaluated. A description of the actual measurement of oxygen uptake is beyond the scope of this section, though further details may be found in all exercise or work physiology textbooks (e.g., Astrand & Rodahl, 1986; Wilmore & Costill, 1994; McArdle, Katch & Katch, 1991). Portable gas analyzers designed with field use in mind are now widely available. Most are lightweight units attached to the patients in a harness. The units typically contain oxygen and carbon dioxide analyzers, a sampling pump, barometric sensors, and battery power supply. The carried units either contain a data logger to store the respiratory data or they have a transmitter that conveys the data in near real time to a nearby personal computer. Although these portable analyzers greatly ease the process of performing indirect calorimetry on personnel in a work setting, they are expensive to purchase and run, require expert knowledge to be deployed effectively, and are not always compatible with rugged field use.

Wilson & Corlett (1995) have proposed a work-intensity classification system based on the  $\text{VO}_2$ . Tasks that use a  $\text{VO}_2$  of more than 2 liters per minute ( $\text{l}\cdot\text{min}^{-1}$ ) are classified as “extremely heavy” and those using  $1.5\text{--}2.0\text{ l}\cdot\text{min}^{-1}$  as “very heavy.” Tasks using  $1.0\text{--}1.5\text{ l}\cdot\text{min}^{-1}$  are considered “heavy,”  $0.5\text{--}1.0\text{ l}\cdot\text{min}^{-1}$  as “moderate,” and less than  $0.5$  as “light.” However, although this classification system categorizes jobs according to their aerobic stress, it fails to take into account the age, sex, and physical fitness of the employee. These factors collectively will determine the strain on the individual employees.

Wilson & Corlett (1995) also provide examples of oxygen uptakes during a number of occupational categories. Assembly work, driving, and office work are reported to have typical oxygen uptakes of  $0.3\text{--}0.6\text{ l}\cdot\text{min}^{-1}$ . Nursing, catering, and light manufacturing have typical oxygen uptakes of  $0.6\text{--}1.0\text{ l}\cdot\text{min}^{-1}$ . Heavy cleaning and manufacturing are cited as having oxygen uptakes of  $0.8\text{--}1.5\text{ l}\cdot\text{min}^{-1}$ ; heavy industrial work, heavy gardening, and agriculture have oxygen uptake of  $1.5\text{--}2.0\text{ l}\cdot\text{min}^{-1}$ ; and firefighting, manual work in forestry, and mining have oxygen uptake of  $2.0\text{--}3.0\text{ l}\cdot\text{min}^{-1}$ .

All activities (leisure and work) and the intensity of the activities can also be classified according to their metabolic equivalent. One metabolic equivalent (MET) equates to resting metabolic rate, which in turn approximates a  $\text{VO}_2$  of  $3.5\text{ ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ . In their textbook, Wilmore & Costill (1994, p. 523) cite the MET values typically associated with a number of occupational tasks. For example, sitting at a desk is allocated 1.5, bricklaying and plastering 3.5, and digging 7.5. However, it should be understood that these figures are mean values and they fail to take into account variations in the rate of work or individual variations in efficiency.

The American College of Sports Medicine (ACSM) Position Stand (American College of Sports Medicine, 1998) classifies physical activity intensity based on physical activity lasting up to 60 minutes in METs by age category. The figures are presented in Table 3.1.

**Table 3.1 ACSM's Classification of physical activity intensity, based on physical activity lasting up to 60 minutes**

Intensity	Endurance-type activity							Resistance-type exercise
	Relative Intensity			Absolute intensity (METs) in healthy adults (age in years)				Relative Intensity
	$\text{VO}_2$ (%) heart rate reserve (%)	Maximum Heart Rate (%)	RPE†	Young (20–39 yr)	Middle-aged (40–54 yr)	Old (65–79 yr)	Very Old (80 + yr)	Maximal voluntary contraction (%)
Very Light	<20	<35	<10	<2.4	<2.0	<1.6	≤1.0	<30
Light	20–39	35–54	10–11	2.4–4.7	2.0–3.8	1.6–3.1	1.1–1.9	30–49
Moderate	40–59	55–69	12–13	4.8–7.1	4.0–5.8	3.2–4.7	2.0–2.9	50–69
Hard	60–64	70–89	14–16	7.2–10.1	6.0–6.4	4.8–6.7	3.0–4.25	70–84
Very Hard	≥55	≥90	17–19	≥10.2	≥8.5	26.8	24.25	≥85
Maximum	100	100	20	12.0	10.0	8.0	5.0	100

\* Based on 8–12 repetitions for persons under age 50–60 years and 10–15 repetitions for persons aged 50–60 years and older.

– Borg rating of Perceived Exertion 6–20 scale (Borg, 1982) (24).

– Maximal values are mean values achieved during maximal exercise by healthy adults. Absolute intensity (METs) values are approximate mean values for men. Mean values for women are approximately 1–2 METs lower than those for men;  $\text{VO}_2$  = oxygen uptake reserve.

– Adapted from and reprinted with permission from U.S. Department of Health and Human Services: Physical Activity and Health: A Report of the Surgeon General. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, 1996:33.

We have measured mean  $\dot{V}O_2$  values from 2 to 3 l.min<sup>-1</sup> for a number of tasks performed by British Army personnel and values from 2.7 to 3.1 l.min<sup>-1</sup> in Royal Navy personnel. The highest values in army personnel were recorded in a group of Infantry assaulting an enemy position (2.92 l.min<sup>-1</sup>), among Royal Engineers bridge-building, Royal Artillery personnel loading and firing artillery ammunition, and Royal Armoured Corps changing tank tracks (all around 2.2-2.3 l.min<sup>-1</sup>) (Rayson, 1998). An investigation of ship-board firefighting performed by naval personnel found drum carrying to be the most aerobically demanding task, requiring a mean of 3.1 l.min<sup>-1</sup> (Bilzon, Scarpello, Smith, Ravenhill, & Rayson), albeit for short periods of several minutes only.

The oxygen requirement can also be related to the individual employee's maximal aerobic power ( $\dot{V}O_{2max}$ ). To expose the relevance of different levels of fitness (which are strongly age and gender related) to the cardiovascular strain on an individual employee, take an example of employee A whose  $\dot{V}O_{2max}$  is 4 l.min<sup>-1</sup> (typical of a young male soldier) and employee B whose  $\dot{V}O_{2max}$  is 2 l.min<sup>-1</sup> (typical of an older female soldier). To perform a task requiring a  $\dot{V}O_2$  of 1 l.min<sup>-1</sup> requires only 25%  $\dot{V}O_{2max}$  for soldier A, whereas it requires 50%  $\dot{V}O_{2max}$  for soldier B. It is thus the percentage of an employee's  $\dot{V}O_{2max}$  that a task demands, together with its duration that determines whether an employee can perform the job.

Astrand and Rodahl (1986) along with many other authors have estimated that the energy demands of an 8-hour day should not exceed 30 to 40% of  $\dot{V}O_{2max}$ . Around 50%  $\dot{V}O_{2max}$  may be sustainable by fit individuals for up to 2 hours, 75% for up to one hour, and 100% for several minutes. Knowing the duration of the task is as important as quantifying the intensity. Returning to our example, although an intensity of 25%  $\dot{V}O_{2max}$  should be tolerable for an 8-hour day for soldier A, an intensity of 50%  $\dot{V}O_{2max}$  would be unsustainable for soldier B.

Another important consideration that is often overlooked is the specificity of HR and  $\dot{V}O_2$  data to a particular mode of activity. Too often in the published literature, authors mix data from different modes of activity (e.g., the  $\dot{V}O_2$  during a manual handling task expressed as a % $\dot{V}O_{2max}$  from treadmill running). Data from Petrofsky & Lind (1978) comparing  $\dot{V}O_{2max}$  of cycling and materials handling and more recent data of our own comparing the two ostensibly similar activities of running and marching (Rayson, Bell, Davies & Rhodes-James, 1995) show that the relationships between HR,  $\dot{V}O_2$ , and activity duration are activity specific and should not be used interchangeably. For example, it would be dangerous to assume that because a person's  $\dot{V}O_{2max}$  on the treadmill is 4 l.min<sup>-1</sup> he could sustain a  $\dot{V}O_2$  of 1.2–1.6 l.min<sup>-1</sup> (30–40% $\dot{V}O_{2max}$ ) all day performing material handling tasks. The  $\dot{V}O_{2max}$  for material handling tasks (or more strictly the  $\dot{V}O_{2peak}$ ) and hence the maximum sustainable level of performance for this type of work would most probably be considerably lower.

**Heart Rate**—Unlike  $\dot{V}O_2$ , which is a measure of cardiovascular stress, heart rate (HR) is a measure of cardiovascular strain. It is a composite index reflecting the psychological and thermal loads as well as the physical demands of the activity, and therefore any interpretation of HR data must consider these other influencing factors. As with  $\dot{V}O_2$ , during moderate-intensity dynamic exercise, HR increases over the first few minutes until a steady state or plateau is reached. Steady-state HR is linearly related to workload and  $\dot{V}O_2$ , and therefore HR may, under certain circumstances, be used as a surrogate for  $\dot{V}O_2$ . However, this should only be done if the relationship between HR and  $\dot{V}O_2$  is known for the employees under investigation during a similar mode of activity. The

gradients of the HR versus V02 and HR versus workload lines depend on fitness as well as the mode of activity—the fitter the employee, the shallower the gradient is.

HR is relatively simple to measure using, for example, over-the-counter heart rate monitors comprising a chest strap that detects and transmits the heart-rate signal, and a wrist monitor that receives and stores the data. HR monitors can be programmed to store data at predefined intervals (e.g., every 5, 30, or 60 seconds) and may be worn by employees throughout the day, with little interference. Given that HR is an unspecific measure of occupational task demands, it is all the more important to monitor and record peripheral information such as temperature and humidity, clothing, and so forth to supplement HR data. If activity data are collected simultaneously, peak and mean HR can be calculated for different tasks performed during the day.

These data can also be used to estimate the intensity of the tasks or job, in terms of the proportion of maximum heart rate (HRmax) ( $220 - \text{age}$ ) or heart rate reserve ( $HRR = (HR_{\text{max}} - HR_{\text{rest}})$ ), both of which can be measured or estimated, and in terms of cardiovascular strain. Wilson & Corlett (1995) suggest that HR during prolonged work of up to 90 beats per minute is considered indicative of light strain, 90–110 as moderate, 110–130 as heavy, 130–150 as very heavy, and 150–170 as extremely heavy, but these classifications do not consider aging. The ACSM Position Stand (ACSM, 1998) classifies physical activity intensity, based on physical activity lasting up to 60 minutes as %HRR, as presented in Table 3.1. %HRR is calculated as  $\frac{(HR - HR_{\text{rest}})}{HRR} \times 100$ .

If measuring working HR is not possible, it might still be possible to measure recovery HR. Recovery HR depends on the HR during work and the fitness of the employee. The quicker the recovery period, the fitter the employee is. According to Brouha's method (Brouha, 1960), the employee is seated immediately after activity and the HR is measured for 3 minutes. The number of heart beats during the first, second, and third minutes is counted:

1. If  $HR_1 - HR_3 \geq 10$ , or if HR1, HR2, and HR3 are all below 90, then recovery is normal.
2. If the average of HR1 over a number of recordings is  $\leq 110$ , and  $HR_1 - HR_3 \geq 10$ , the workload is not excessive.
3. If  $HR_1 - HR_3 < 10$ , and if  $HR_3 > 90$ , then recovery is inadequate.

In Worked Example 2 that follows, we provide an illustration of how HR data collected in the field can be combined with laboratory measurements of oxygen uptake to estimate the cardiovascular demands of a task. Minimum acceptable standards of aerobic fitness can be established by this method. The data are from a recent Military project that we conducted to define the physical demands of rural patrolling in soldiers.

**Body Temperature**—Body temperature is normally maintained around 37°C, but during strenuous exercise or in extreme conditions of heat or cold, temperatures can fluctuate outside of these values. Body temperature reflects a fine balance between heat gain and heat loss. During exercise, most of the energy the body produces is converted to heat. If the body's heat production exceeds its loss, core temperature rises. Heat loss can take place via the processes of radiation, conduction, convection, and evaporation. But if the ambient temperature is greater than the body's temperature, heat may be gained, not lost, by radiation, conduction, and convection. Similarly, exposure to the sun's



## Worked Example 2

We measured HR over a 5-day exercise using Polar HR monitors, downloading the data at the end of each working day. By Wilson & Corlett's criteria (1995), HR during most activities in the majority of soldiers equated to a 'moderate' workload, indicating that the workload was sustainable without undue physical fatigue. However, HR in a minority of soldiers equated to a "heavy" workload that would result in fatigue and impaired performance over time. On average 24% (122 min), 8% (41 min), and 2% (6 min) of time on patrol, were spent above 60%, 70%, and 80%HR<sub>max</sub>, respectively. Eight percent of patrol time (41 minutes) was spent at intensities of work above the soldier's anaerobic threshold, indicating the importance of anaerobic, as well as aerobic fitness to patrolling. Two of the five days were significantly more demanding than were the other three. Soldiers within the platoon fulfilling a particular function and carrying particular items of equipment had a lower HR than the remaining soldiers performing all other roles. Surprisingly, these individuals were found to be carrying above-average loads. On further investigation, high levels of aerobic fitness were found to account for this apparent anomaly. Without the additional data on weight of loads and fitness, an erroneous conclusion could easily have been made.

We also used the same HR data to estimate the energy cost of the patrols using the HR versus V<sub>O</sub>2 relationship that we had calculated from a simulated loaded march on a treadmill (i.e., a near-identical mode of activity to patrolling) in all participants. Steady-state HR and V<sub>O</sub>2 values from the last 30 seconds of each 3-minute workload of the treadmill test were regressed to form a linear equation in the form of  $V_{O2} = m \times HR + c$ . Resting HR values were taken as the minimum HR recorded during sleep. This value was plotted against a theoretical resting oxygen uptake value of 3.5 ml.kg.<sup>-1</sup>.min.<sup>-1</sup> (1 MET). R<sup>2</sup> values for the equations, ranged from 0.88 to 0.99 percent, indicating that the equations fitted the data well (i.e., 88% to 99% of the variation in V<sub>O</sub>2 could be accounted for by HR).

Using the daily HR data and the individually determined HR versus V<sub>O</sub>2 relationship, we estimated V<sub>O</sub>2 for each soldier throughout the day's patrolling activities. The overall mean oxygen uptake of the 5 days of rural patrols was 1.10 l.min.<sup>-1</sup> (SD 0.46) or 15.0 ml.kg.<sup>-1</sup>.min.<sup>-1</sup> (SD 6.4), which equates to a "heavy" workload by Wilson & Corlett's classification (1995) and a "light" intensity by ACSM's classification for the young (20–39 years) (ACSM, 1998). The mean time in minutes and the proportion of the total time spent at workloads above 40%, 50%, 60%, 70%, and 80% of individual V<sub>O</sub>2<sub>max</sub> were also calculated. The soldiers worked at an overall mean %V<sub>O</sub>2<sub>max</sub> of 33% (SD 6), with mean values per day ranging from 26% (SD 6) to 37% (SD 7). On the most physically demanding day, 50% of soldiers performed at a work intensity above 40% V<sub>O</sub>2<sub>max</sub> (the suggested maximum sustainable intensity of work). Occasional fairly brief spurts of activity that were very demanding were recorded. Peak oxygen uptakes during the most demanding 10 minutes of the exercise averaged just over 2 l.min.<sup>-1</sup>. Peak oxygen uptakes during the most demanding 60 minutes of the patrol averaged approximately 1.60 l.min.<sup>-1</sup>. On the basis of these last figures (60 min at 1.6 l.min.<sup>-1</sup>) we proposed minimum standards of aerobic fitness of 3.2 l.min.<sup>-1</sup>, based on the premise that approximately 50%V<sub>O</sub>2<sub>peak</sub> could be sustained for 1 hour.

radiation may result in heat gain not loss. During exercise the primary route of heat loss is evaporation of sweat, though this too may be impaired by clothing, high humidity, and/or dehydration.

Prolonged heavy sweating, the body's normal primary heat loss mechanism during exercise in ambient conditions can also result in fluid loss if it is not replaced at an equivalent rate. Fluid loss leads to reduced blood volume, and reduced blood volume leads to reduced cardiac output, curtailing our ability to continue working as efficiently and to dissipate heat as effectively. Heart rate and body temperature become elevated during exercise when more than 2% of body mass (fluid) is lost. Exercise and work performance decline by approximately 10% at 2% body mass lost and 25% at 4% body mass lost (Saltin & Costill, 1988, in Wilmore & Costill, 1994).

Strenuous physical work in hot environments imposes large stresses and strains on the employee. It is important as a minimum to monitor ambient conditions during a job analysis; and where a significant thermal load is suspected, measurements of deep body (core) and skin temperature should be made.

Ambient conditions are normally monitored by measuring the wet bulb globe temperature (WBGT)—a weighted composite index of wet bulb temperature, radiant heat, and dry bulb temperature. The index can be used to describe conditions in which work is performed. It can also be used to predict the risk of heat injury for exercise at different intensities. Further details about thermal considerations may be obtained from any thermal physiology textbook (e.g., Parsons, 1993) and the International Standards Organization (ISO) series of publications (e.g., ISO, 1993).

Body temperature can be assessed in numerous ways, including oral, aural, esophageal, and rectal for deep body temperature, or on the skin for peripheral temperature. Due to the variation in the temperature of different tissues, mean temperature is sometimes used. Mean temperature is calculated as a weighted average of skin and internal temperatures. For example skin thermistors may be placed on the arm ( $T_a$ ), trunk ( $T_t$ ), leg ( $T_l$ ), and head ( $T_h$ ) and skin temperature would be calculated as

$$T_{skin} = (0.1 \times T_a) + (0.6 \times T_t) + (0.2 \times T_l) + (0.1 \times T_h)$$

The constants in the equation represent the proportion of the total skin area represented by each region. Once core temperature ( $T_c$ ) has been measured, mean body temperature can be calculated as

$$T_{body} = (0.4 \times T_{skin}) + (0.6 \times T_c)$$

Worked Example 3 describes a relatively noninvasive novel technique for measuring deep body or core temperature that we have used recently as part of a job analysis. To our knowledge, the device is not commercially available currently, but it offers exciting potential. The purpose of including this physiological technique in our job analysis was to ascertain the presence and extent of any hypothermic and hyperthermic strain in soldiers carrying out rural patrols.

**Stress hormones**—Measuring the concentration of stress hormones in the body provides another avenue for assessing physiological (and psychological) strain as high levels of strain may be reflected in changes in endocrine function. Blood levels of cortisol may increase, while thyroxine and testosterone levels may be depressed (Wilmore & Costill, 1994). Resting *blood* levels of epinephrine and norepinephrine may also be elevated resulting in raised heart rate and blood pressure.

However, collecting of these blood hormones is an invasive procedure, and measurement is complex and expensive. Further, unless baseline data are obtained under normal conditions in the

### Worked Example 3

We used a Thermal Monitoring System originally designed for measuring core body temperature in deep-sea divers (Mekjavic, Tomsic, Gider, Golder, & Tipton, 1996; Tipton, Franks & Golder, 1997). The system comprised a radio pill, data logger, and two temperature sensors. The nonrecoverable pill (in itself, a major advance relative to previous recoverable pills), contains a blocking oscillator near-field transmitter powered by a battery, all of which is encapsulated in medical grade epoxy used in surgical implants.

The radio pill is ingested. An AM receiver monitors the radio pill emissions and the pulses sampled are stored in the random access memory of the logger. The pulse frequency is converted to temperature values ( $^{\circ}\text{C}$ ) based on a calibration equation previously derived for each pill. The logger is programmed and the collected data retrieved by connecting it to a personal computer using Mini-Mitter software.

The length of time the pill remains in the intestine will vary depending on the intestinal motility, which is influenced by the diet, but typically, the pill will stay in the body for 1 to 3 days. Few studies have used a radio pill to measure core body temperature, therefore less is known about the normal expected ranges. However, it is unlikely that the temperature of the gastrointestinal tract will differ from other sites used to measure core body temperature (rectal, esophageal, tympanic) by more than  $0.5^{\circ}\text{C}$ .

Core body temperatures between  $35^{\circ}\text{C}$  and  $39^{\circ}\text{C}$  are unlikely to cause any serious health problems if the soldier is able to regulate at these temperatures and is not at the extremes for long periods. "Normal" body temperatures throughout the day in cool ambient conditions (experienced during this study) and low exercise intensities are between  $36^{\circ}\text{C}$  and  $38^{\circ}\text{C}$ .

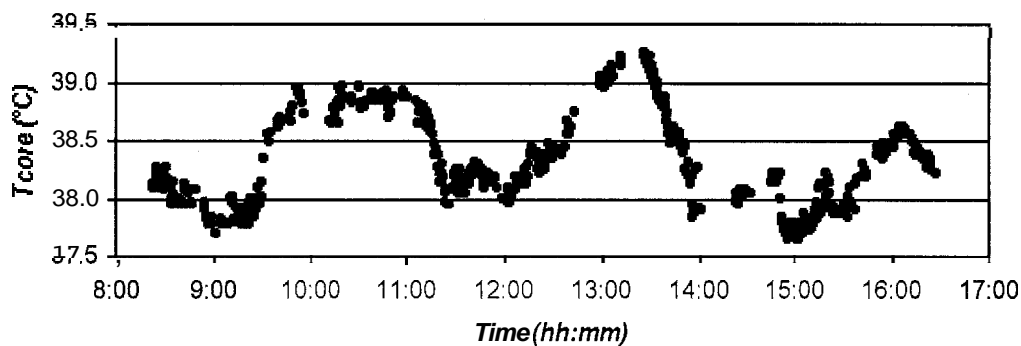
In women, body temperature will vary with the menstrual cycle. Resting and exercising core body temperatures are higher during the midluteal compared with the late follicular phase (Kolka & Stephenson, 1997). In addition, women appear to have a lower heat tolerance, a higher sweating threshold and a lower sweating capacity than men (Fox, Lofstedt, Woodward, Eriksson, & Werkstrom, 1969). Fitness will also affect temperature regulation and may mask any gender differences. The core body temperature at which an individual reaches steady state during exercise depends on the relative rather than the absolute exercise intensity. Therefore a fitter individual will have a lower core body temperature for a given work load (Saltin & Hermansen 1966).

An example core body temperature trace from our study is provided in Figure 3.2. The relatively high starting temperature of  $38.2^{\circ}\text{C}$  at approximately 08.20 indicates that the soldier had exercised, had been in a hot environment, or possibly had consumed a hot drink before the measurements started (the pill would have been in the stomach at this stage and may have been in physical contact with the hot fluid). The first increase in temperature occurred between 09.30 and 11.00 and corresponded with an HR averaging  $124\text{b}\cdot\text{min}^{-1}$ .

The plateau in the temperature from 10.00 to 11.00 indicates that the soldier was able to thermoregulate with a steady state temperature of 38.7°C. Temperature decreased when the exercise ceased at 11.00. The next increase in temperature corresponded with a mean HR of 138 b.min<sup>-1</sup>, between 12.00 and 13.20. The temperature showed no signs of reaching a plateau, indicating that the soldier was unable to thermoregulate at this exercise intensity. However, when the soldier stopped at 13.20, temperature declined rapidly.

Collection of core body temperature data using ingested telemetry pills proved socially acceptable among the soldiers and viable in the field and elicited interesting data. This technology provides a useful measurement technique where rectal temperature measurement is not possible. Three of our nine soldiers (33%) exhibited core body temperatures in excess of 39°C, the generally accepted upper limit for safe operations. There was no evidence found of these soldiers' ability to thermoregulate at this temperature (no plateau in the trace could be detected), providing some cause for concern. No soldiers exhibited temperatures below 35°C, the generally accepted lower limit for safe operations.

#### 7 Mar: Deep Body Temperature



*Figure 3.2 Trace of deep body temperature throughout a day's patrolling activity in a soldier, measured via an ingested telemetry pill. Reprinted by permission from ACSM, 7998 Classification of Physical Activity Intensity Position Stand. 30.6, 978.*

specific sample of Participants involved in the job analysis, the findings are difficult to interpret, as variability both within and between people can be considerable. These procedures are therefore not normally applicable to conducting a Physical Demands Analysis, unless the demands are extremely high and a state of overtraining or exhaustion in the work force is suspected.

We have found salivary cortisol to be a potentially useful marker of stress as it can be obtained by noninvasive methods. During acute exercise or stress, only a minor increase in cortisol is found, but during chronic exercise or stress, levels can increase by 150 percent in 30 minutes. Cortisol levels show a circadian pattern, generally peaking in the morning around breakfast time, reaching the lowest values during late morning, and then rising gradually throughout the afternoon, evening, and night (Nieman, 1996), so obtaining baseline data in all participants on a "normal" nonworking day is important. However, investigating salivary cortisol still requires access to biochemistry labo-

ratories for analysis. In addition, interpreting the data remains challenging as threshold values indicating elevation and a high degree of strain appear to be quite variable between individuals, and accepted normative values have yet to be established.

**Global Positioning System and energy prediction equations**—Recently, we have been experimenting and had reasonable success with using Portable Global Positioning Systems (GPS) to collect positional data and time of soldiers in the field. During our recent study on patrolling, each soldier carried a GPS in the top of his rucksack while deployed. The post-processed coordinates were fed into an AXIS digital mapping system, which forms part of a concept demonstrator program. Vertical distances were calculated by overlaying the GPS data onto contoured maps. For the Physical Demands Analysis, we calculated the mean distance covered per patrol to be 4.134 km, mean speed to be 1.4 km/h, with 63 meters of vertical ascent and 23 meters of vertical descent (i.e., the noncircular patrol route ended at an altitude 40 meters higher than at the start).

From this information we estimated the energy cost of the patrols, using a formula developed by Pandolf, Givoni & Goldman (1977). This approach provides an alternative method to estimating the aerobic demand of an activity in the field without measuring any physiological data (e.g., HR and  $\text{VO}_2$ ).

$$\text{Metabolic cost of walking (watts)} = 1.5W + [2(W + L)] \times \left[\left(\frac{L}{W}\right)^2 + T(W + L)\right] \times [(1.5V^2 + 0.35VG)]$$

Where :

$W$  = body mass (kg)

$L$  = load mass (kg)

$T$  = terrain factor

$V$  = velocity (m/s)

$G$  = gradient (%)

The Pandolf equation has been compared with observed data by a number of authors. It has been found to predict slightly high for standing with loads and low for walking at slow speeds on both grade and level walking (Pimental & Pandolf 1979; Pimental, Shapiro & Pandolf, 1982). Duggan & Ramsay (1987) reported predictions averaged 3% too high, but generally reported good agreement with measured values while walking at 6 km.hr<sup>-1</sup> on the level with and without a 21 kg load. However, the equation is not valid for downhill walking, which limits its application. Epstein, Stroschein & Pandolf (1987) developed an equation for predicting the metabolic cost of running with and without backpack loads.

## Biomechanical Approach

The biomechanical approach scrutinizes the forces exerted on and by the body during work. It is often used to analyze postures and the support and movement of loads and is therefore particularly useful in assessing most material handling tasks. The body is viewed as a system of levers and joints in which the levers are rotated around the joint by the action of skeletal muscles. The mus-

cles attach close to the joint, allowing a small contractile distance of the muscle to be transformed into large movements of the distal end of the lever. The mechanical advantage of the loads at the distal end of the lever over the muscles result in the generation of large muscle forces to overcome relatively small loads (Ayoub & Mital, 1989).

Biomechanics is a useful tool in our repertoire for conducting a Physical Demands Analysis. It is particularly useful for comparative studies in which different conditions or methods for performing a task are compared, rather than for precise quantification of workload. Some of the more useful biomechanical methods available to us include posture recordings, measurements of force and maximum voluntary contractions (MVC), measurements of muscular activity via electromyography (EMG), and estimates from biomechanical analyses.

**Posture**—The recording and analysis of posture during task performance may be useful for several reasons, including descriptive purposes and to identify any health and safety issues. Depending on the accuracy required, images from photographs or videos may suffice for recording postures and calculating approximate body segment angles. Two synchronised images taken at 90° to each other provide additional information, but there remain inaccuracies caused by parallax. For a biomechanical analysis greater accuracy of data is ideally required, but for assessing posture in relation to discomfort, strain, stability, or force exertions, simple observational methods are adequate (Wilson & Corlett, 1995).

The OWAS method is a posture coding system designed in Finland for industrial use (Finnish Institute of Occupational Health, 1992). Postures are observed and recorded using a recording sheet and each posture is assessed for acceptability using an assessment sheet. A six-figure OWAS code is generated: the first three figures record the posture, the fourth figure indicates the force, and the fifth and sixth figures indicate the task being performed. The recommended method is to glance at the work at predefined sampling intervals and then to look away and record the data. From these samples, estimates of the proportion of time spent in different postures and the forces exerted can be estimated. Although the precision of the data is low, the method enables rapid identification of the major inadequate postures during force exertion. McAtamney & Corlett (1993) have developed an analogous procedure to assess exposure of employees to the risk of upper limb disorder called Rapid Upper Limb Assessment (RULA).

To improve accuracy over direct observational methods of posture, goniometers can be used to measure angles between body segments or between body segments and the vertical. Where spinal measures are made, Corlett suggests using the goniometer at L3 to L5 to estimate the angle at the lumbar-sacrum junction, and on the lower part of the thoracic spine to estimate spinal angle.

Technological developments have led to the production and widespread availability of computerised electronic devices for measuring spinal motion (Marras, Ferguson & Simon, 1990). The devices are secured to the body at the chest and hips. Velocity and acceleration as well as motion are calculated by the software.

There are now a number of commercial systems on the market that use multiple video cameras and sophisticated computer software to track markers placed at strategic points on subjects performing their work activities. Although these advances ease the process of capturing work activity, it remains a long and tedious process to extract and analyse the data. Estimates of analysis time of up to 10 times that of recorded time are not unusual. Mainly for this reason, simpler more direct measurements are often favored in a work setting.

**Force measurement**—Given the logarithmic relationship between the amount of time a force can be sustained and the proportion of MVC that this force represents, it is often useful in tasks that involve a significant force component to measure both the forces involved and the MVC in the relevant muscle groups and postures in employees.

Measuring the loads and forces involved in work tasks can often be done quite simply by weighing any objects to be lifted and carried and by inserting a force transducer between the employee and the object in the case of pushing and pulling tasks. In the case of tasks that involve movement, initial forces **will** often be greater than sustained forces since they **will** reflect the extra force required to overcome inertia to start the object in motion. The greater the acceleration is, the greater the initial force **will** be. The speed and acceleration of movements by employees therefore have an impact on the forces exerted, and efforts should be made to control these parameters. Employees often perform tasks with additional vigor when performing for the investigator's benefit. Investigators should be aware of this tendency and take steps to counter it.

If we calculate the percentage that the measured force represents as a proportion of an employee's MVC for that muscle group and posture, and we know its duration and frequency, we can judge its acceptability by reference to accepted threshold values. We could also estimate the maximum duration for which a given force could be sustained in **all** employees for whom we have an MVC, but that information is less useful.

**An** alternative approach that is commonly adopted in developing occupational fitness standards is to measure the static or dynamic strength of relevant muscle groups and to relate these to task performance using regression equations (the criterion validity approach). This approach has been adopted by many authors usually with reasonable success (e.g., Poulsen, 1970; Sharp et al., 1980; Pytel & Kamon, 1981; Ayoub et al., 1982; Teves, Wright & Vogel, 1985; Nottrodt & Celentano, 1987; Beckett & Hodgdon, 1987; Dueker, Ritchie, Knox & Rose, 1994; Rayson, Holliman & Belyavin, 2000).

In these studies, static strength tests that were strongly associated with lift performance included upright pull, back, arm, shoulder, and leg strength. The most strongly correlated dynamic strength and power tests with lift performance included the Incremental Lift Machine test, bench press, vertical and broad jump, and isokinetic back extension and isokinetic lift power. The highest correlation coefficients for men and women separately were reported by Pytel & Kamon (1981) between isokinetic strength scores and lift scores ( $r = 0.87-0.96$ ). However, more typically  $r$ -values of 0.3–0.5 were reported for single sex data (e.g., Myers, Gebhardt, Crump, & Fleishman, 1984; Teves, Wright & Vogel, 1985). Usually the  $r$ -values were higher for men than for women (e.g., Myers et al., 1984; Teves et al., 1985).

To improve predictive capability many studies build multiple regression models combining several strength and anthropometric variables in the prediction equation.  $R^2$  values ranged from 0.33 and 0.11 for men and women respectively (Teves et al., 1985) up to 0.94 in a pooled gender sample (Pytel and Kamon, 1981). **A** number of studies pooled the male and female data apparently without first checking the validity of this technique in their population—this practice should be avoided.

**Electromyography**—EMG is a technique for measuring electrical activity of the muscle. It can be used to assess the involvement of a muscle group in a particular movement or task, the extent of the muscle involvement, and the state of fatigue of the muscle (Hagberg, 1981). Needle electrodes inserted into the muscle is the preferred technology, but usually in an occupational setting nonin-

vasive surface electrodes are used over the belly of the muscle. The signal is amplified and normally recorded as an electronic file for later analysis.

The presence or absence of electrical activity indicates the active involvement or otherwise of a particular muscle group. Further, the consistent exponential relationship between EMG activity and force during both static and dynamic activities (Hagberg, 1981) enables the user to establish the extent of involvement of muscle groups and estimate the forces involved. To perform this **type** of analysis, EMG response to an employee's MVC in exactly the same posture also needs to be measured. Without maximal data on the employees under investigation, there is no means of anchoring and interpreting the response, other than in qualitative terms.

These data can be especially useful in isolating the role of specific muscle groups in complex activities such as loaded marching. For example, under relatively light backpack loads the electrical activity of the erector spinae (back extensors) is lower than with no load, while under heavy load, the muscle group is clearly more active than with no load (Knapik, Harman & Reynolds, 1996). The gastrocnemius shows similar increases in EMG activity with load, indicating an increased demand on this muscle group.

Fatigue of muscle can also be detected and estimated via EMG. Fatigue exhibits itself as an increase in the amplitude in the low frequency range of EMG activity and also as a shift in the frequencies toward the lower end of the spectrum.

**Biomechanical models**—Once forces and torques have been calculated, we can start to make comparisons between tasks and within tasks under various conditions. We can also assess the feasibility and safety of tasks by comparing the data with appropriate population norms and with occupational safety and health guidelines. Below, we review the National Institute for Occupational Safety and Health (NIOSH) equation—a widely used tool—and its use in conducting a Physical Demands Analysis. To cover further detail on these topics, refer to Ayoub & Mital (1989) for an overview of 2D and 3D, static and dynamic, biomechanical models; to the UK Defence Standards 00–25 (Ministry of Defence, 1998); and to Snook & Ciriello (1991) for Military and industrial population norms, respectively.

NIOSH published guidelines for manual lifting in 1981 (National Institute of Occupational Safety and Health, 1981), which provided an empirical method for computing a load limit for lifting. The guide considered the epidemiology of musculoskeletal injury, and set biomechanical, physiological, and psychological limits. Its application was limited to two-handed symmetrical, smooth lifting directly in front of the body using handles. **Six** different factors are used to determine the action limit (AL)—the object load, the horizontal distance between ankles and hands, the vertical distance between hands and floor, the vertical travel distance, the frequency, and the duration of the lifting task. These **six** parameters are entered into the NIOSH equation.

The AL is defined as the load that can be safely handled by 75 percent and 99 percent of women and men, respectively. Thus lifting loads that fall below the AL are considered safe. The maximum permissible limit (MPL), which is three times the AL, is the load that can be safely handled by only 25% of men and virtually no women. Loads that fall between the AL and the MPL pose an increased risk for some workers. Loads above the MPL pose a significant risk for many workers.

In 1991 the NIOSH equation was revised (Waters, Putz-Anderson, Garg & Fine, 1993) to encompass a broader range of tasks, and alterations were made to the various criteria. **Six** coeffi-



cients were used to reduce the load constant to compensate for characteristics of the lift that were different from the standard conditions.

For a Physical Demands Analysis, the NIOSH equation serves two uses. First, it can be used to provide an index of acceptability for each task that is performed by the work force. Second, by calculating a ratio between the weight of lift and the actual weight of lift, we can obtain an index of task difficulty.

In Worked Example 4 an illustration is provided of how we used the NIOSH equation and the AL and MPL to classify and interpret material handling tasks performed by all Career Employment Groups in the British Army.

### Worked Example 4

As part of a Physical Demands Analysis that we conducted on occupations within the British Army (Rayson, 1998), we submitted the data relating to all of the material handling tasks to the NIOSH (1981) equation. We found that only 12% of tasks were “acceptable”—falling below AL, while 67% required redesign—falling between AL and MPL. The remaining 21% of tasks exceeded MPL—these were considered unsafe to perform under any circumstances by the NIOSH criteria. Among the nine tasks that fell into this last category, four exceeded the AL by fourfold.

## Psychophysical Approach

The psychophysical approach assumes that both biomechanical and physiological stresses impinge on an employee performing any task and that these stresses are integrated and combined and can be assessed as an objective measure of acceptable demand rate of repetitive work or perceived stress. There are several psychophysical methods that may be useful during a Physical Demands Analysis, including obtaining ratings of perceived exertion (RPE) (Borg, 1982, 1985, 1998) and perceived effort (Fleishman, Gebhardt & Hogan, 1984), and using the Body Map for evaluating body part discomfort (Wilson & Corlett, 1995). For a full discussion of measurement of psychological demand refer to the textbook *A Guide to Manual Materials Handling* by Mital, Nicholson & Ayoub (1993).

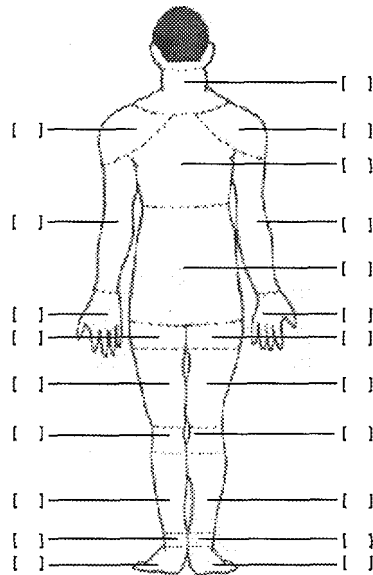
**Borg’s Rating of Perceived Exertion**—The Borg Scale (Borg, 1998) is based on the linear relationship that exists between workload, heart rate, and perceived exertion. In the original scale, which spans 6 (no exertion at all) through 20 (maximal exertion), the ratings approximately correspond to heart rate divided by 10. Some authors also use an alternative nonlinear 10-point Category Ratio Scale (the CR-10), but this version appears to be less widely used.

The scale is described to the employee before the activity commences and is presented to the employee at fixed time intervals or during particular target activities to obtain an RPE. The RPEs may provide additional subjective information to accompanying physiological measurements or they may be used instead of HR measurements when the latter are not feasible for whatever reason. Similar Visual Analogue Scales are used to rate pain.

**Fleishman's Perceived Effort Index**—Building on the work of Borg and co-workers in Sweden, Fleishman developed a 7-point effort scale intended for use in the workplace. On the early version of the scale, each number was anchored by verbal descriptions (e.g., 1 = very, very light; 4 = somewhat hard; 7 = very, very hard). Fleishman found that both job-experienced and inexperienced men and women were able to make fine distinctions between tasks at all ranges of effort. For example, lifting and carrying objects weighing 85 to 100 pounds (mean rating **6.6**) and laying railroad tracks (mean rating **6.3**) were rated as the most physically demanding tasks (Hogan & Fleishman, 1979). A correlation of 0.81 between the metabolic cost and the mean ratings for **30** tasks demonstrated the strength of the relationship.

The scale was later modified by replacing the adjectives with behavioral task anchors of high, medium, and low effort (e.g., operate a jackhammer (mean 5.91), perform light welding (mean 3.27), and operate a calculator (mean 1.08) (Fleishman, Gebhardt, & Hogan, 1984). The reliability and validity of the perceived effort index led the authors to conclude that it could be substituted for actual physiological measurement of work across a wide variety of physically demanding tasks. Herein lies the usefulness and potential for this index in performing a Physical Demands Analysis.

**Body Map**—Wilson and Corlett's Body Map (Wilson & Corlett, 1995) is a useful tool for evaluating both the location and severity of body part discomfort. As displayed in Figure 3.3, the body is divided into segments—the actual size and clustering of the segments can be modified to suit the application. The scale can be used in a number of ways. In its simplest form, employees can be shown the scale during or after performing particular work tasks, and they can indicate in which parts of the body they are experiencing discomfort. If more time is available for measurement, severity of discomfort can also be rated using a 5-point paper scale anchored at the 0 and 5 points by “no discomfort” and “extreme discomfort.” Care should be taken when summarizing the data—calculating mean ratings may not be appropriate since mean values will mask interesting individual differences that might be linked to body size, particular items of equipment used, age, or gender. We have found it useful to use two body maps simultaneously, representing the front and back of the body. Lower back, hamstring, and calf discomfort, for example can then be more precisely reported.



**Figure 3.3** Copyright 1995, From Wilson and Corlett's body map by J. R. Wilson & E. N. Corlett. Reproduced by permission of Taylor & Francis, Inc., <http://www.routledge-ny.com>

## Summary

In this chapter, we have presented and discussed the relative merits of a number of approaches and techniques for conducting a job analysis. These have included some industrial psychological techniques such as observation, questionnaire, and interview for identifying the most physically demanding tasks. Once the critical and most frequently performed tasks have been identified, various physiological, biomechanical, and psychophysical techniques can be deployed to quantify the stress and strain associated with these tasks. These have been described and worked examples provided of our own experience of their application.

The physiological techniques reviewed include measurement of oxygen uptake to estimate energy expenditure, heart rate to assess cardiovascular strain, body temperature to investigate thermal strain, hormones to quantify stress levels, and Global Positioning System to track movement and to estimate energy expenditure. The biomechanical techniques presented comprise posture analysis to describe body position and to identify any health and safety issues, force measurement to quantify the extent of the forces exerted, electromyography to assess muscle involvement and fatigue, and the use of biomechanical models to make comparisons between tasks and within tasks. The psychophysical techniques encompass subjective rating tools such as Borg's Rating of Perceived Exertion, Fleishman's Perceived Effort Index, and Wilson and Corlett's Body Map.

The selection of approach and techniques by the investigator will depend on many factors, including the job or task under investigation, the resources and time available, and the expertise of

the investigation team. Generally, a multidisciplinary approach performed by a multiskilled team is preferred because it is more likely to elicit a complete and balanced output. Time should be taken by the investigating team to reflect and to discuss.

In conclusion, conducting a job analysis to identify and quantify the most physically demanding key tasks in a job is a complex process that requires considerable investment of time, money, and effort. Good science and good judgment are required in equal measure. The result is a solid foundation on which to base selection, training, and retention fitness criteria. The payback will be increased productivity through improved operational effectiveness and reduced injury.

---

## References

1. Aghazadeh, F., & Ayoub, M. M. (1985). A comparison of dynamic and static-strength models for prediction of lifting capacity. *Ergonomics*, 28, (10), 1409–1417.
2. American College of Sports Medicine. (1998). The recommended quantity and quality of exercise for developing and maintaining cardiorespiratory and muscular fitness, and flexibility in healthy adults. Position Stand. *Medicine & Science in Sports & Exercise*, 30, (6), 978.
3. Annett, J., Duncan, K. D., Stammers, R. B., & Gray, M. J. (1971). *Task Analysis*. Training Information Paper No 6. HMSO, London.
4. Astrand, P. O., & Rodahl, K. (1986). *Textbook of work physiology: Physiological bases of exercise*. McGraw Hill International Editions.
5. Ayoub, M. M., Denardo, J. D., Smith, J. L., Bethea, N. J., Lambert, B. A., Alley, L. R., & Duran, B. S. (1982). *Establishing physical criteria for assigning personnel to Air Force jobs*. (Final Report. **Air** Force Office of Scientific Research, Contract No. F49620-79-C-0006). Institute for Ergonomics Research, Texas Tech University, Lubbock, TX.
6. Ayoub, M. M., & Mital, A. (1989). *Manual materials handling*. London: Taylor & Francis.
7. Beckett, M. B., & Hodgdon, J. A. (1987). *Lifting and carrying capacities relative to physical fitness measures*. Technical Report 87-26, Naval Health Research Center, Bethesda, MD.
8. Bilzon, J. L. J., Scarpello, E. G., Smith, C. V., Ravenhill, N. A., & Rayson, M. P. (submitted for publication). *Characterisation of the metabolic demands of simulated shipboard Royal Navy fire-fighting tasks*.
9. Borg, G. A. V. (1982). Psycho-physical bases of perceived exertion. *Medicine and Science in Sports and Exercise*, 14, (5), 377–381.
10. Borg, G. (1985). *An introduction to Borg's RPE scale*. Ithaca, New York: Movement Publications.
11. Borg, G. (1998). *Borg's perceived exertion and pain scales*. Champaign, IL: Human Kinetics.
12. Brouha, L. (1960). *Physiology in industry*. Oxford: Pergamon.
13. Bruce, R. A. (1971). Exercise testing of patients with coronary heart disease. *Annals of Clinical Research*, 3, 323.
14. Chaffin, D. B. (1974). Human strength capability and low-back pain. *Journal of Occupational Medicine*, 16, (4), 248–254.
15. Courville, J., Vezina, & Messing, K. (1991). Comparison of the work activity of two mechanics: A woman and a man. *International Journal of Industrial Ergonomics*, 7, 163–174.

16. Cunningham, J. W., Boese, R. R., Neeb, R. W., & Pass, J. J. (1983). Systematically derived work dimensions: Factor analysis of the occupational analysis inventory. *Journal of Applied Psychology*, 68, 232–252.
17. Desmond, R. E., & Weiss, D. J. (1973). Superior estimation of abilities required in jobs. *Journal of Vocational Behavior*, 7, 13–29.
18. Desmond, R. E., & Weiss, D. J. (1975). Worker estimation of ability requirements of their jobs. *Journal of Vocational Behavior*, 7, 13–29.
19. Drenth, P. J. D., Thierry, H., & de Wolff, C. J. (1998). *Handbook of work and organizational psychology* (second edition). Volume 3: Personnel Psychology. Hove, UK: Psychology Press.
20. Dueker, J. A., Ritchie, S. M., Knox, T. J., & Rose, S. J. (1994). Isokinetic strength testing and employment. *Journal of Occupational Medicine*, 36, (1), 42–48.
21. Duggan, A., & Ramsay, D. A. (1987). A respiratory assessment of PLCE (Inf) combat order with and without chest strap. Army Personnel Research Establishment, Working Paper 29/87, Farnborough.
22. Dwyer, O. W., Prien, E. P., & Burke, T. L. (1987). Making physical fitness standards job related. *Parks and Recreation*, 22, (2), 30–32.
23. Epstein Y., Stroschein, L. A., & Pandolf, K. B. (1987). Predicting metabolic cost of running with and without backpack loads. *European Journal of Applied Physiology*, 56, 495–500.
24. Ferguson, G. A., & Takane, Y. (1989). *Statistical analysis in psychology and education*. McGraw Hill International Editions.
25. Fine, S. A., & Cronshaw, S. F. (1999). *Functional job analysis: a foundation for human resource management*. New York: John Wiley & Sons.
26. Fine, S. A., & Wiley, W. W. (1971). An introduction to functional job analysis. In E. A. Fleishman & A. R. Bass (Eds.), *Studies in personnel and industrial psychology*. Homewood, IL: Dorsey Press.
27. Finnish Institute of Occupational Health. (1992). *OWAS, a method for the evaluation of postural load during work*. Publication Office, Topeliuksenkatu 41 aA, SF 00250 Helsinki, Finland.
28. Flanagan, J. C. (1954). The critical incidents technique. *Psychological Bulletin*, 51, 327–358.
29. Fleishman, E. A., Gebhardt, D. L., & Hogan, J. C. (1984). The measurement of effort. *Ergonomics*, 27, (9), 947–954.
30. Fleishman, E. A., & Quintance, M. K. (1984). *Taxonomies of human performance*. Orlando, FL: Academic Press.
31. Fox, R. H., Lofstedt, B. E., Woodward, P. M., Eriksson, E., & Werkstrom, B. (1969). Comparison of thermoregulatory function in men and women. *Journal of Applied Physiology*, 26, 444–53.
32. Frykman, P. N., & Harman, E. A. (1995). Anthropometric correlates of maximal locomotion speed under heavy backpack loads. *Medicine and Science in Sports and Exercise*, 27 (Suppl.), 5.
33. Hagberg, M. (1981). An evaluation of local muscular load and fatigue by electromyography. *Arbete och Halsa*, 24, Solna, Sweden.
34. Harwood, G. E., Rayson, M. P., & Nevill, A. M. (1999). Fitness, performance and risk of injury in British Army officer cadets. *Military Medicine*, 164, (6), 428–434.
35. Hogan, J. C., & Fleishman, E. A. (1979). An index of the physical effort required in human task performance. *Journal of Applied Psychology*, 64, 197–204.
36. ISO DIS 11399. (1993). *Ergonomics of the thermal environment: Principles and application of international standards*. Geneva, International Standards Organisation.

37. Jones, B. H., Bovee, M. W., & Knapik, J. J. (1992). The association between body composition, physical fitness and injuries among male and female army trainees. In B. M. Marriott & B. J. Scott (Eds.), *Body composition and physical performance*. Washington, DC: National Academy Press, 141–173.
38. Kelly, G. A. (1955). *The psychology of personal constructs*. New York: Norton.
39. Kirwan, B., & Ainsworth, L. K. (1992). *A guide to task analysis*. London: Taylor & Francis.
40. Knapik, J. J., Harman, E., & Reynolds, K. (1996). Load carriage using packs: A review of physiological, biomechanical and medical aspects. *Applied Ergonomics*, 27, (3), 207–216.
41. Kolka, M. A. (1992). Temperature regulation in women. *Med Exer Nutr Health*, 1, 201–207.
42. Kolka, M. A., & Stephenson, L. A. (1997). Effect of luteal phase elevation in core temperature on forearm blood flow during exercise. *Journal of Applied Physiology*, 82, 1079–83.
43. Kuorinka, I., Jonsson, B., Kilbom, A., Vinterberg, H., Biering-Sorensen, F., Andersson, G., & Jorgensen, K. (1987). Standardized Nordic questionnaires for the analysis of musculoskeletal symptoms. *Applied Ergonomics*, 18, 233–237.
44. Lopez, F. M. (1986). *The thresholds traits analysis*. Port Washington, New York: Lopez & Associates.
45. McArdle, W. D., Katch, F. I., & Katch, V. L. (1991). *Exercise physiology: Energy, nutrition and human performance*. London: Lea & Febiger.
46. McAtamney, L., & Corlett, E. N. (1993). RULA: A survey method for the investigation of work-related upper limb disorders. *Applied Ergonomics*, 24, 91–99.
47. McCormick, E. J., Jeanneret, P. R., & Mecham, R. C. (1972). A study of job characteristics and job dimensions as based on the Position Analysis Questionnaire. *Journal of Applied Psychology*, 56, 347–368.
48. McDaniel, J. W. (1998). Design requirements for human strength Asking the right questions. In Kuman S. (Ed.), *Advances in Occupational Ergonomics and Safety 2*, IOS Press, Amsterdam, Netherlands, 345–348.
49. Marras, W. S., Ferguson, S. A., & Simon, S. R. (1990). Three dimensional dynamic motor performance of the normal trunk. *International Journal of Industrial Ergonomics*, 6, 211–214.
50. Mekjavic, I. B., Tomsic, M., Gider, F., Golder, F. StC., & Tipton, M. J. (1996). Thermal monitoring system. In Y. Shapiro, D. S. Moran, & Y. Epstein (Eds.), *Environmental Ergonomics—Recent progress and new frontiers* (pp. 421–433). Freund Publishing House Ltd., Tel Aviv, Israel.
51. Ministry of Defence. (1998). *Defence Standards Issue 3*, June 1998. Ministry of Defence (Procurement Executive). Directorate of Standardization. UK.
52. Mital, A. (1986). Subjective estimates of load carriage in confined and open spaces. In Karwowski, W. (Ed.), *Trends in ergonomics/human factors III*, Amsterdam: North Holland, pp. 827–833.
53. Mital, A., & Fard, H. F. (1986). Psychophysical and physiological responses to lifting symmetrical and asymmetrical loads symmetrically and asymmetrically. *Ergonomics*, 29, 1263–1272.
54. Myers, D. C., Gebhardt, D. L., Crump, C. E., & Fleishman, E. A. (1984). *Validation of the military entrance physical strength capacity test*. Technical Report 6 10, USARIEM, MA, USA.
55. National Institute of Occupational Safety & Health. (1981). *Work practices guide for manual lifting*. Department of Health and Human Services. (NIOSH) Publication No. 81–122, USA.
56. Nieman, D. C. (1996). Effect of long-term training on the immune system and on resistance to infectious diseases. In R. J. Maughan & S. M. Shirreffs (Eds.), *Biochemistry of exercise* (pp. 383–398). Champaign, IL: Human Kinetics.

57. Nottrodt, J. W., & Celentano, E. J. (1987). Development of predictive selection and placement tests for personnel evaluation. *Applied Ergonomics*, 18, (4), 279–288.
58. Pandolf, K. B., Givoni, B., & Goldman, R. F. (1977). Predicting energy expenditure with loads while standing or walking very slowly. *Journal of Applied Physiology*, 43, (4), 577–581.
59. Parsons, K. C. (1993). *Human thermal environments*. London: Taylor & Francis.
60. Petrofsky, J. S., & Lind, A. R. (1978). Comparison of metabolic and ventilatory responses of men to various lifting tasks and bicycle ergometry. *Journal of Applied Physiology: Respiratory, Environmental and Exercise Physiology*, 45, 60–63.
61. Petrofsky, J. S. & Lind, A. R. (1978). Comparison of metabolic and ventilatory responses of men to various lifting tasks and bicycle ergometry. *Journal of Applied Physiology: Respiratory, Environmental and Exercise Physiology*, 45, 60–63.
62. Pimental, N. A., & Pandolf, K. B. (1979). Energy expenditure while standing or walking slowly uphill or downhill with loads. *Ergonomics*, 22, (8), 963–973.
63. Pimental, N. A., Shapiro, Y., & Pandolf, K. B. (1982). Comparison of uphill or downhill walking and concentric and eccentric cycling. *Ergonomics*, 25, (5), 373–380.
64. Poulsen, E. (1970). *Prediction of maximum loads in lifting from measurements of muscular strength*. In Communications from the Danish National Association for Infantile Paralysis.
65. Pugh, L. G. C. E., Corbett, J. L., & Johnson, R. H. (1967). Rectal temperatures, weight losses in marathon running. *Journal of Applied Physiology*, 23, 1079–83.
66. Pytel, J. L., & Kamon, E. (1981). Dynamic strength test as a predictor for maximal and acceptable lifting. *Ergonomics*, 24, (9), 663–672.
67. Rayson, M. P. (1998). The development of physical selection procedures. Phase 1: job analysis. In M. A. Hanson (Ed.), *Contemporary Ergonomics* (pp. 393–393). London: Taylor & Francis.
68. Rayson, M. P., Bell, D. G., Davies, A., & Rhodes-James, E. (1995). Heart rate and oxygen uptake relationship: A comparison of loaded marching and running in women. *European Journal of Applied Physiology*, 71, 405–408.
69. Rayson, M. P., Holliman, D., & Belyavin, A. (2000). Development of physical selection procedures for the British Army. Phase 2: Relationship between physical performance tests and criterion tasks. *Ergonomics*, 43, (1), 73–105.
70. Rayson, M. P., Pynn, H., Rothwell, A., & Nevill, A. (2000). The development of physical selection procedures for the British Army. Phase 3: Validation. In P. T. McCabe, M. A. Hanson, & S. A. Robertson (Eds.), *Contemporary Ergonomics*, Taylor and Francis, London, 140–144.
71. Rice, V. J., & Sharp, M. A. (1994). Prediction of performance on two stretcher-carry tasks. *Work*, 4, (3), 201–210.
72. Ridd, J. E. (1985). Spatial restraints and intra-abdominal pressure. *Ergonomics*, 28, 149–166.
73. Saltin, B., & Costill, D. L. (1988). In Wilmore JH, & Costill DL (Eds.) (1994). *Physiology of Sport and Exercise* (p. 365). Champaign, IL: Human Kinetics.
74. Saltin, B., & Hermansen, L. (1966). Esophageal, rectal and muscle temperature during exercise. *Journal of Applied Physiology*, 21, 1757–62.
75. Saville and Holdsworth Limited (1989). *Work Profiling System*. London: Saville and Holdsworth Ltd.

76. Sharp, D. S., Wright, J. E., Vogel, J. A., Patton, J. F., Daniel, W. L., Knapik, J., & Korval, D. M. (1980). *Screening for physical capacity in the U.S. Army: An analysis of measures predictive of strength and stamina*. Technical Report T8/80, United States Army Research Institute of Environmental Medicine, Natick, USA.
77. Sinclair, M. A. (1975). Questionnaire Design. *Applied Ergonomics*, 6, 73–80.
78. Snook, S. H. (1978a). The design of manual handling tasks. *Ergonomics*, 21, 963–985.
79. Snook, S. H. (1978b). *The design of manual handling tasks*. Paper presented at Ergonomics Society Lecture, Cranfield Institute of Technology, Bedfordshire, England.
80. Snook, S. H., & Ciriello, V. M. (1974). The effects of heat stress on manual handling tasks. *American Industrial Hygiene Association Journal*, 28, 322–329.
81. Snook, S. H., & Ciriello, V. M. (1991). The design of manual handling tasks: Revised tables of maximum acceptable weights and forces. *Ergonomics*, 34, 1197–1213.
82. Stevenson, J. M., Bryant, T., Greenhorn, D., Smith, T., Deakin, J., & Surgenor, B. (1990). The effect of lifting protocol on comparisons with isoinertial lifting performance. *Ergonomics*, 33, (12), 1455–1469.
83. Teves, M. A., Wright, J. E., & Vogel, J. A. (1985). *Performance on selected candidate screening test procedures before and after Army basic and advanced individual training*. Technical Report T13/85, United States Army Research Institute of Environmental Medicine, Natick, USA.
84. Tipton, M. J., Franks, C. M., Golder, FStC. (1997). *Thermal monitoring of operational dives phase 1 final report: Equipment development and evaluation*. Robens Institute, University of Surrey, Report No. RI97/HAP/003.
85. Waters, T. R., Putz-Anderson, V., Garg, A., & Fine, L. J. (1993). Revised NIOSH equation for the design and evaluation of manual lifting tasks. *Ergonomics*, 36, (7), 749–776.
86. Wilmore, J. H., & Costill, D. L. (1994). *Physiology of Sport and Exercise*. Human Kinetics, Champaign, IL.
87. Wilson, J. R., & Corlett, E. N. (Eds.) (1995). *Evaluation of human work: A practical ergonomics methodology*. London: Taylor and Francis.







# Chapter 4

## Types of Physical Performance Tests

**Andrew S. Jackson, P.E.D.**

Department of Health and Human Performance  
University of Houston  
Houston, TX

---

### Abstract

This chapter reviews the two general types of tests used to evaluate a person's ability to do physically demanding work basic ability tests and work sample tests. The basic ability tests reviewed include aerobic fitness, body composition, strength, muscle endurance, and flexibility. The common laboratory and field tests were reviewed and evaluated. Work sample tests are designed to duplicate occupational tasks. The weaknesses and strengths of work sample tests are discussed. Research confirms that physically demanding work sample test performance largely depends on aerobic fitness, body composition, and strength to varying degrees. Although important, flexibility, balance, and agility are less likely to be related to physically demanding tasks. This chapter reviews these data. When basic ability tests are highly correlated with work sample tests, one test administration option is to replace the work sample test with a basic ability test or combination of basic ability tests. The final section of the chapter is a brief review of the role of aerobic fitness, strength, and body composition on health and injury

---

### Introduction

The types of tests used to evaluate one's fitness to perform physically demanding work tasks can be categorized into two general types: physical ability tests and work sample tests. Physical ability tests measure the basic fitness components of aerobic capacity, body composition, strength, muscular endurance, and flexibility. Physical ability tests not only evaluate a person's capacity to do

demanding work tasks but also their physical fitness. In contrast, work sample tests just evaluate a person's ability to perform a work task or combinations of tasks.

This chapter is an overview of basic ability tests that measure aerobic capacity, body composition, muscular endurance, strength, and flexibility. Work sample tests along with their strengths and limitations are discussed. Next, the relationship between basic ability tests and work sample tests is examined. Although the primary objective of this State-of-the-Art Report (SOAR) is to examine the role of physical fitness on performing physically demanding jobs, fitness also has a health promotion component. The final section of the chapter examines the role of fitness on health and risk of injury.

---

## Physical Ability Tests

This section includes the basic physical ability tests that measure aerobic fitness, body composition, strength, muscular endurance, and flexibility as well as an overview of the nature and types of tests used to measure these physical abilities.

### Aerobic Fitness

Aerobic fitness, the maximal volume of oxygen one can consume during exhausting exercise ( $\text{VO}_{2\text{max}}$ ), depends on several factors including efficient lungs, heart, and blood vessels; the quality and quantity of blood (red blood count and volume); and the cellular components that help the body use oxygen during exercise. Because a person's ability to use oxygen during exhaustive work depends on these factors,  $\text{VO}_{2\text{max}}$  is an accepted test of aerobic fitness and an indicator of subsequent exercise capacity (ACSM, 1990; ACSM, 1991). Åstrand and Rodahl (Åstrand & Rodahl, 1970) consider it to be the best index of physical fitness:

*During prolonged heavy physical work, a person's performance capacity depends largely upon his ability to take up, transport, and deliver oxygen to working muscle. Subsequently, the maximal oxygen uptake is probably the best laboratory measure of a person's physical fitness, providing the definition of physical fitness is restricted to the person's capacity for prolonged heavy work. (Åstrand & Rodahl, 1970, p. 314)*

There are several different ways to categorize and contrast aerobic fitness tests. These are—

1. *laboratory tests or field tests,*
2. *maximal or submaximal tests, and*
3. *tests that do not involve exercise, that is, nonexercise tests (Baumgartner & Jackson, 1999).*

A discussion of these tests follows.



$$\text{Bruce Treadmill Test Healthy Subjects (Foster et al., 1964), } R = 0.97 \quad (2)$$

$$VO_{2max} \text{ (ml/kg/min)} = 17.50 - (0.30 \times T) + (0.297 \times T^2) - (0.0077 \times T^3)$$

The Bruce protocol is a nonlinear equation developed with both healthy subjects and cardiac patients (Foster et al., 1984). The provided Bruce equation is for persons free of coronary heart disease. The  $VO_{2max}$  of heart patients is 4.2 ml/kg/min lower than healthy patients for a given maximum treadmill time.

Submaximal Aerobic Fitness Laboratory Tests—Exercising to  $VO_{2max}$  is physically exhausting, time-consuming, expensive, and requires medical supervision when testing high-risk subjects. The laboratory submaximal tests provide a less accurate but easier and safer method of estimating aerobic fitness. The measurement objective of submaximal tests is to define the slope of a person's heart-rate response to exercise and use the slope to estimate  $VO_{2max}$  from submaximal parameters. The three exercise physiological principles that are the foundation of submaximal tests are as follows—

- \* Heart rate (i.e., pulse rate) increases in direct proportion to the oxygen used during aerobic exercise.
- $VO_{2max}$  is reached at maximum heart rate.
- A less fit person will have a higher heart rate at any submaximal level than someone who is more aerobically fit.

Oxygen uptake ( $VO_2$ ) at any level of exercise is the product of cardiac output and the difference in the oxygen content of the arterial and venous blood. Cardiac output, the volume of blood pumped with each heart beat, is the product of heart rate and stroke volume. Stroke volume increases early in exercise and stabilizes at about 45 percent of  $VO_{2max}$ . The testing goal of estimating  $VO_{2max}$  from submaximal power output is to measure heart rate, between 45 and 70 percent (= 115 to 150 b/min) of a person's  $VO_{2max}$ . Below 45 percent of  $VO_{2max}$ , stroke volume has not leveled off, whereas at about 70 percent  $VO_{2max}$ , exercise is likely shifting from aerobic to anaerobic. Submaximal tests use a submaximal aerobic power output. Singlestage and multistage models estimate  $VO_{2max}$  from submaximal power output and exercise heart rate (Baumgartner & Jackson, 1999; Ross & Jackson, 1990).

The multistage exercise test requires that heart rate and power output be measured at two or more submaximal levels (Golding, Meyers, & Sinning, 1989). These data points project to maximal heart rate, which estimates aerobic fitness. The multistage model is the procedure used for the popular YMCA adult fitness test (Golding et al., 1989). The YMCA test uses a cycle ergometer following a branching protocol to regulate power output for each 3-minute stage. The goal of the test is to obtain at least two submaximal heart rates between 115 and 150 b/min.  $VO_{2max}$  is estimated by plotting the linear increase in exercise heart rate associated with increases in power output. Connecting the two points defines the linear power output and the heart rate slope. The line defined by the slope is extended to maximum heart rate, estimated by  $220 - \text{age}$ , and  $VO_{2max}$  is estimated by dropping the line down to the power output scale expressed in a metric of absolute  $VO_2$  (ml/min).

The Singlestage Exercise Test Model is both simpler to use and slightly more accurate than the Multistage Model (Mahar, Jackson, & Ross, 1985). It was initially popularized by the Astrand-Rhyming nomogram (Astrand & Rodahl, 1970; Astrand & Rodahl, 1986; Astrand & Ryhming,

1954). Research (Jackson et al., 1990) showed that the correlation between  $\dot{V}O_{2\max}$  estimated from the single stage equation and Astrand-Rhyming nomogram was 0.99. Equation 3 is the mathematical representation of the Astrand-Rhyming nomogram (Baumgartner & Jackson, 1999; Jackson et al., 1990; Ross & Jackson, 1990).

$$\text{Single-Stage Submaximal } \dot{V}O_{2\max} \text{ Equation} \quad (3)$$

$$\dot{V}O_{2\max} = \dot{V}O_2 SM \times \frac{220 - \text{Age} - k}{SM \cdot HR - k}$$

Power output can be regulated with either the cycle ergometer or treadmill. The  $\dot{V}O_2$  SM term in Equation 3 is the  $\dot{V}O_2$  at submaximal exercise. Standard treadmill or cycle ergometer energy cost equations estimate power output (ACSM, 1991; Baumgartner & Jackson, 1999). The term SM HR is the submaximal heart rate at  $\dot{V}O_2$  SM, and k is the constant of 61 for men and 73 for women. The constants represent the intercept of the heart rate and  $\dot{V}O_2$  relationship, that is, a heart rate for a  $\dot{V}O_2$  of 0.

The errors associated with submaximal tests are the measurement of exercise heart rate, the use of the term 220 – age as an accurate representation of “true maximum heart rate,” and estimating submaximal  $\dot{V}O_2$  when it is not directly measured. The standard error of representing maximum heart rate by 220 – age is about  $\pm 10$  beats/min (ACSM, 1991). Although this does affect accuracy somewhat, it is not a major source of error. What introduces major systematic prediction errors are conditions that affect the heart rate response to exercise. A major source of inaccuracy is drugs such as beta blockers that lower both exercise and maximum heart rate. Submaximal tests are not suitable for subjects taking drugs that alter heart rate response to exercise. These drugs lower both exercise and maximum heart rate thus producing an overestimate of true  $\dot{V}O_{2\max}$ . Another source of error is submaximal  $\dot{V}O_2$ . It is rarely measured during an exercise test. Standard equations are available to estimate  $\dot{V}O_2$  from cycle ergometer (Astrand & Rodahl, 1986; Åstrand & Ryhming, 1954) and treadmill power output (ACSM, 1991; Ross & Jackson, 1990). Research (Jackson et al., 1990; Ross & Jackson, 1986) shows these estimates are a major source of error.

**Maximal Distance Run Field Tests**—Running performance has been shown to be related with  $\dot{V}O_{2\max}$ . When performed properly, running tests provide valid assessments of aerobic fitness, but are not as accurate as the regression equations derived for maximum treadmill protocols in which the speed and elevation are strictly regulated. The most common run tests involve jogging and/or walking distances ranging from 1 to 3 miles, or traveling as far as possible in 12 minutes.

Dr. Kenneth Cooper was one of the first to popularize distance run/walk tests. The goal of his research was to provide a field test to assess the aerobic fitness of U.S. Air Force personnel (Cooper, 1968). His sample consisted of 115 airmen who ranged considerably in age (17 to 52 years), weight (114 to 270 pounds), and aerobic fitness (28 to 60 ml/kg/min). The subjects first completed the distance run/walk test of the miles covered in 12 minutes. On the next day the subjects completed a maximal treadmill test in which  $\dot{V}O_{2\max}$  was measured by indirect calorimetry. Cooper found a very high correlation ( $r = 0.90$ ) between distance covered in 12 minutes and measured  $\dot{V}O_{2\max}$ . He published the following regression equation with a function to estimate distance traveled in miles from  $\dot{V}O_{2\max}$  (ml/kg/min)—

$$\text{Cooper's 12-minute Run/Walk Model} \quad (4)$$

$$\text{Distance (miles)} = 0.3138 = (0.0278 \times \text{VO}_2\text{max ml/kg/min})$$

Cureton and Associates (Cureton, Sloniger, O'Bannon, Black, & McCoormack, 1995) published a comprehensive study relating 1-mile run/walk performance with  $\text{VO}_2\text{max}$ . Their heterogeneous sample consisted of more than 750 men and women who ranged in age from 8 to 25 years. The goal of the study was to develop a generalized regression equation that provided valid estimates of aerobic fitness for youth and adults of both genders.

These researchers found that the relationship between  $\text{VO}_2\text{max}$  and mile run/walk time was not linear and that gender, age, and body mass index (BMI) accounted for aerobic fitness variance. The multiple correlation of the generalized equation was 0.72, and the standard error of estimate was 4.8 ml/kg/min. Equation 5 gives the generalized 1-mile run/walk equation. The term  $T$  is mile run/walk time in minutes, and  $G$  is gender coded: female = 0, male = 1; and  $\text{BMI}$  is body mass index.

$$\text{Cureton's 1-mile Run/Walk Generalized Equation} \quad (5)$$

$$\text{VO}_2\text{max (ml/kg/min)} = 108.94 - (8.41 \times T) + (0.34 \times T^2) + (0.21 \times \text{Age} \times G) - (0.84 \times \text{BMI})$$

Distance run tests are the most commonly used field test of aerobic fitness. Note that a timed distance run test is a maximal test. It has all the attendant risks of a maximal test with the added risk of being unsupervised. Maximal distance run tests are suitable only for young people in good condition without significant cardiovascular disease risk factors.

**Submaximal Field Test — The Rockport Walk Test**—A limitation of a laboratory submaximal test is the need for a cycle ergometer or treadmill to regulate power output. The Rockport Walk Test (Kline, Porcari & Hintermeister, 1987) provides a means of estimating  $\text{VO}_2\text{max}$  from heart rate response to walking speed. Track and heart-rate monitoring equipment are needed to administer the test. The Rockport Test involves walking as fast as possible for 1 mile, and then measuring exercise heart rate immediately after the walk. The data needed to estimate  $\text{VO}_2\text{max}$  include the following—

- Weight measured in pounds
- Mile walk time
- Exercise heart rate (beats/min) measured immediately at the end of the walk
- Age measured to the last year
- Gender coded, female = 0, and male = 1.

Multiple regression equations were developed to estimate  $\text{VO}_2\text{max}$  (ml/kg/min) from these variables. A general equation was developed for men and women. The Rockport Walk Test is shown in Equation 6 where  $W$  is body weight in pounds,  $T$  is mile run time in minutes,  $HR$  is exercise heart rate, and  $G$  is gender, female = 0 and male = 1.



$$\begin{aligned} &\text{Rockport WalkTest ( } R=0.88, \text{ SEE } = 5 \text{ ml/kg/min)} \\ &VO_2\text{max (ml/kg/min)} = \\ &132.85 - (0.39 \times \text{Age}) - (0.08 \times W) - (3.26 \times T) - (0.16 \times HR) + (6.32 \times G) \end{aligned} \quad (6)$$

An assumption of the Rockport WalkTest is that exercise heart rate is at a steady state. This is best assured by walking at a brisk steady pace. Since the Rockport Test is a submaximal test, factors that affect exercise heart rate reduce its accuracy. For example, someone taking beta blocker drugs will have a lower than normal exercise heart rate. This will lead to a systematic overestimate of  $VO_2\text{max}$ . These prediction errors can be quite large. Another important consideration is pacing.

**Nonexercise Aerobic Fitness Tests—** $VO_2\text{max}$  can be estimated without testing subjects (Jackson et al., 1990). Aerobic fitness can be estimated with reasonable accuracy from a person's age, gender, body composition, and self-report level of aerobic exercise. A very large database of NASA/Johnson Space Center (Houston, TX) employees was used to develop the nonexercise models.  $VO_2\text{max}$  was measured by indirect calorimetry. Before being tested, the employees rated their physical activity (Figure 4.1) during the previous month. Multiple regression was used to estimate  $VO_2\text{max}$  from exercise rating in combination with age, gender, and a body composition parameter consisting of either percent body fat or body mass index.

The original research (Jackson et al., 1990) provided two equations that could be used for men and women. The difference in the equations was the body composition variable. The most accurate equation ( $R = 0.81$ ,  $SEE = 5.3 \text{ ml/kg/min}$ ) used skinfold-determined percent body fat (Jackson & Pollock, 1978; Jackson, Pollock, & Ward, 1980). The equation that used BMI was slightly less accurate ( $R = 0.78$ ,  $SEE = 5.7 \text{ ml/kg/min}$ ), but more feasible for mass testing. A limitation of these early equations was that the women's sample size was much smaller than that of the men's. Additional studies were published from this database (Jackson et al., 1995; Jackson et al., 1996b) in which the goal was to examine the influence of aging on  $VO_2\text{max}$ . A much larger sample of women was used, providing a means of developing gender specific equations. Following are the male and female equations (Equations 7 to 10), where AR is activity code (0 to 7), % fat is percent body fat, and BMI is body mass index. The BMI equations are published in another source (Baumgartner & Jackson, 1999).

$$\begin{aligned} &\text{Percent Fat Non-exercise Men's Equation ( } R=0.79, \text{ SEE } = 4.9 \text{ ml/kg/min)} \\ &VO_2\text{max (ml/kg/min)} = \\ &47.820 - (0.259 \times \text{Age}) - (0.216 \times \%fat) + (3.275 \times AR) - (0.082 \times \%fat \times AR) \end{aligned} \quad (7)$$

$$\begin{aligned} &\text{Percent Fat Non-exercise Women's Equation ( } R=0.85, \text{ SEE } = 4.4 \text{ ml/kg/min)} \\ &VO_2\text{max (ml/kg/min)} = \\ &45.628 - (0.265 \times \text{Age}) - (0.309 \times \%fat) + (2.175 \times AR) - (0.044 \times \%fat \times AR) \end{aligned} \quad (8)$$

$$\begin{aligned} &\text{BMI Non-exercise Men's Equation ( } R=0.74, \text{ SEE } = 5.4 \text{ ml/kg/min)} \\ &VO_2\text{max (ml/kg/min)} = \\ &55.688 - (0.362 \times \text{Age}) - (0.331 \times BMI) + (4.310 \times AR) - (0.096 \times BMI \times AR) \end{aligned} \quad (9)$$

$$\text{BMI Non-exercise Women's Equation } (R=0.82, SEE=4.7 \text{ ml/kg/min}) \quad (10)$$

$$VO_{2max} \text{ (ml/kg/min)} =$$

$$44.310 - (0.326 \times \text{Age}) - (0.227 \times \text{BMI}) + (4.471 \times \text{AR}) - (0.135 \times \text{BMI} \times \text{AR})$$

Code for Physical Activity
<p>Use the appropriate number (0 to 7) which best describes your general <b>ACTIVITY LEVEL</b> for the <b>PREVIOUS MONTH</b>.</p> <p><i>Do not participate regularly in programmed recreation sport or heavy physical activity.</i></p> <p>0 – Avoid walking or exertion, e.g., always use elevator, drive whenever possible instead of walking.</p> <p>1 – Walk for pleasure, routinely use stairs, occasionally exercise sufficiently to cause heavy breathing or perspiration.</p> <p>2 – 10 to 60 minutes per week.</p> <p>3 – Over one hour per week.</p> <p><i>Participate regularly in heavy physical exercise such as running or jogging, swimming, cycling, rowing, skipping rope, running in place, or engaging in vigorous aerobic activity type exercise such as tennis, basketball or handball.</i></p> <p>4 – Run less than one mile per week or spend less than 30 minutes per week in comparable physical activity</p> <p>5 – Run 1 to 5 miles per week or spend 30 to 60 minutes per week in comparable physical activity</p> <p>6 – Run 5 to 10 miles per week or spend 1 to 3 hours per week in comparable physical activity</p> <p>7 – Run over 10 miles per week or spend over 3 hours per week in comparable physical activity.</p>

**Figure 4.1 Scale for rating level of physical activity. The directions are to select one value that best represents the level of physical activity for the previous month. The scale was developed for use in the Cardio-pulmonary Laboratory, NASA/Johnson Space Center, Houston, Texas. Source: Baumgartner and Jackson (1999).**

The nonexercise tests are especially feasible for mass testing. Since heart rate is not a factor of the nonexercise test, the nonexercise equations are valid for subjects taking heart-rate altering medication (Jackson et al., 1990). With the ease of test administration, one may question the accuracy of them. It has been demonstrated that the nonexercise tests were more accurate than the single stage submaximal test (Jackson et al., 1990), but less accurate than maximal treadmill performance (Baumgartner & Jackson, 1999). The obvious limitation of the nonexercise approach is the use of subjective rating

for their level of exercise. The nonexercise models are especially useful for estimating the  $\text{VO}_{2\text{max}}$  of large numbers of subjects. The BMI models provide a means of obtaining  $\text{VO}_{2\text{max}}$  by self-report. This provides a means of estimating the  $\text{VO}_{2\text{max}}$  of large samples by questionnaire.

Evaluation Standards—Table 4.1 lists the age and gender aerobic fitness standards recommended by the American College of Sports Medicine (Gettman, 1993). These standards reflect the well-established gender difference in  $\text{VO}_{2\text{max}}$  and the decline in  $\text{VO}_{2\text{max}}$  with age (Buskirk & Hodgson, 1987; Jackson et al., 1995; Jackson et al., 1996b).

**Table 4.1 American College of Sports Medicine standards for  $\text{VO}_{2\text{max}}$  ml/kg/min \***

Standard	Age in Years				
	20-29	30-39	40-49	50-59	≥60
<b>Men</b>					
Excellent	≥52	249	247	≥43	241
Good	49-51	46-48	44-46	4 M 2	38-40
Average	42-48	39-45	37-43	33-39	31-37
Fair	39-41	36-38	34-36	30-32	28-30
Poor	538	535	≤33	529	527
<b>Women</b>					
Excellent	243	≥40	238	234	≥34
Good	4W2	37-39	35-37	31-33	31-33
Average	33-39	31-36	29-34	25-30	25-30
Fair	30-32	28-30	26-28	22-24	22-24
Poor	529	527	525	521	≤21

\* from Gettman (1993). Reprinted, by permission, from L. R. Gettman, 1993, American College of Sports Medicine: Resource manual for guidelines for exercise testing and prescription (2<sup>nd</sup> ed., pp. 229-246). Philadelphia: Lea & Febiger

Although the norms in Table 4.1 give aerobic fitness standards for adults, the levels are not suitable for adult health promotion. The research published by Blair and associates (1989) gives the first scientific data defining the aerobic fitness needed for health. It showed that the aerobic fitness health promotion threshold was 32 ml/kg/min for women and 35 ml/kg/min for men (Figure 4.2). The mortality rate of men and women with the lowest level of aerobic fitness was four times higher than the rate of men and women who exceeded these levels. Aerobic fitness declines with age, and the 35 and 32 levels were for men and women at age 45. Table 4.2 lists health promotion aerobic fitness standards adjusted for the age-related decline in aerobic fitness. Also listed is the value needed to have an aerobic power of 35 or 32 ml/kg/min at age 45 years, assuming one maintains his or her current level of exercise and percent body fat. Research shows that changing exercise habits and percent body fat affects the rate that aerobic fitness changes with age (Jackson et al., 1995; Jackson et al., 1996b).

## Body Composition

Suitable levels of body composition are important for health and the capacity to perform work tasks that require a person to move his or her body weight. Following are the methods used to evaluate body composition.

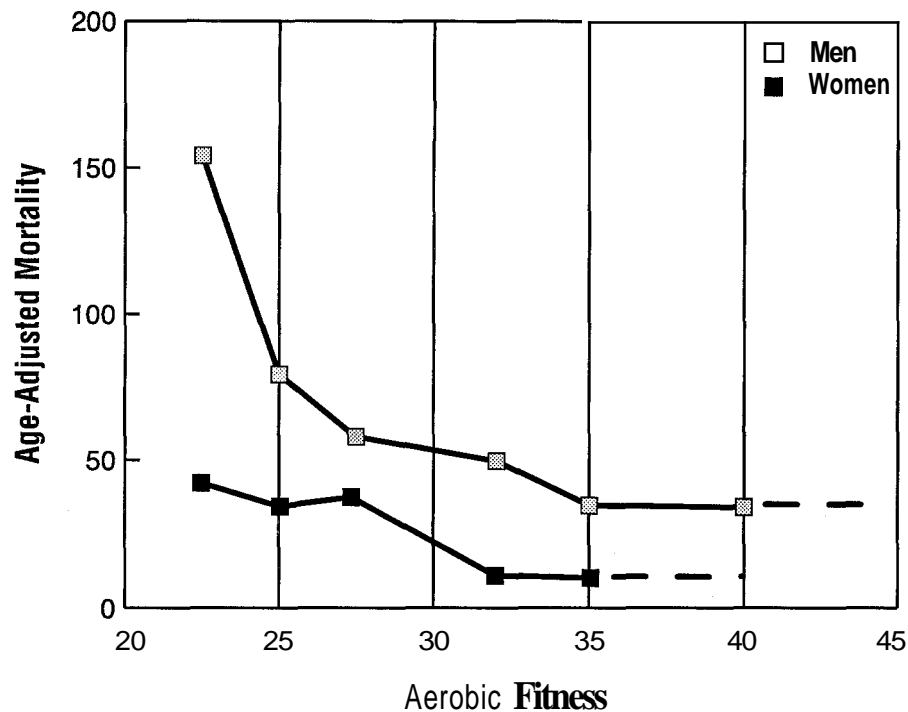


Figure 4.2 Curves of the relationship between aerobic fitness and health show that the curves level off at 32 ml/kg/min for women and 35 ml/kg/min for men. These values define the threshold level of fitness needed for health promotion (i.e., reduced mortality) for 45-year-old men and women. *Journal of American Medical Association*, 1989, 262, pp. 2395–2401. “Copyrighted 1989, American Medical Association”

Table 4.2 Age-adjusted adult aerobic fitness standards for health promotion\* above  $\dot{V}O_{2\max}$  ml/kg/min

Age Group	Men	Women
45 and under	35	32
50	34	31
55	32	29
60	31	28
65 and Over	30	27

\* Standards developed from data (Jackson et al., 1995, Jackson et al., 1996) and personal communication with S. Blair, September 30, 1993. From Baumgartner and Jackson (1999).

**Body Density and Percent Body Fat**—In simple terms, body weight consists of fat weight and fat-free weight. Percent body fat is simply the proportion of total weight that is fat weight. Percent body fat is measured from body density, the ratio of body weight, and body volume. The hydrostatic or underwater weighing method is the most common laboratory method used to measure body composition. Numerous laboratories at universities and medical centers have the equipment for underwater weighing determinations. The measurement objective of hydrostatic weighing is to measure body volume, which is then used with body weight to calculate body density. Percent fat is calculated from body density. A newer, less common method for measuring body volume is with a “bodybox” or body plethysmograph.

The values needed to calculate body volume (BV) are body weight on land ( $W_t$ ), body weight in water ( $W_w$ ), the density of water ( $D_w$ ), and the body’s air component ( $B_a$ ) consisting of resid-

ual volume + 100 ml. The 100 ml value is an estimate of air in the gastrointestinal tract. Equation 11 gives the equation used to measure body volume. Body density (BD) is the ratio of body weight on land and body volume (Equation 12).

$$\begin{array}{l} \text{Body Volume} \\ BV = \frac{W_t - W_w}{D_w} - Ba \end{array} \quad (11)$$

$$\begin{array}{l} \text{Body Density} \\ BD = \frac{W_t}{BV} \end{array} \quad (12)$$

Variation in body density can be caused by air, fat weight, and fat-free weight. The density of air is zero, and the density of fat weight tissue is about 0.90 g/cc. The density of fat-free weight varies from about 1.0 g/cc to as high as 3.0 g/cc. Fat-free weight consists of muscle, blood, bone, and organs. The two-component models for computing percent body fat from body density were based on the assumption that the density of fat tissue was 0.90 g/cc and fat-free weight was 1.10 g/cc. Researchers are starting to question this assumption when computing the percent body fat of children, the elderly, and ethnic groups. This has led to the development of multicomponent models (Heymsfield, 1996; Lohman, 1992).

The first two-component equations developed for converting body density to percent body fat were published by Siri, 1961 and Brozek et al. (Brozek, Grande, & Anderson, 1963). The equations provide nearly identical percent fat values throughout the human range of body fatness. The equations are as follows—

$$\begin{array}{l} \text{Siri Percent Body Fat} \\ \%fat = \frac{495}{BD} - 450 \end{array} \quad (13)$$

$$\begin{array}{l} \text{Brozek Percent Body Fat} \\ \%fat = \frac{457}{BD} - 414 \end{array} \quad (14)$$

There is growing evidence that the Siri and Brozek equations may not be accurate when applied to some ethnic groups. The ethnic differences in body density are believed to be because of fat-free weight differences associated with bone mineral content (Sinning, 1996; VanLoan, 1996). Cross-sectional data (Vickery et al., 1988) show that the mean body density of African-American men (1.075 g/cc) was significantly higher than white men (1.065 g/cc). They also found that the mean of the sum of seven skinfolds was not different, suggesting that the difference in the relationship of skinfolds to body density for white men versus African-American men was due to variability in the composition of fat-free weight.

Sinning (1996) suggests that the high-bone mineral content of African-Americans results in a fat-free density higher than the value of 1.1 g/cm assumed by the Brozek and Siri equations and recommends that an equation published by Schutte et al. (Schutte, 1984) be used to convert body density to percent for African-American men. The equation is as follows

$$\%fat = \frac{437.4}{BD} - 392.8$$

The Siri, Brozek, and Schutte methods of estimating percent body fat from body density have lost their status as the “gold standard” for assessing body composition. Each is based on the two-component model that assumes that the density of fat tissue is 0.9 g/cc and the body’s average density of fat-free weight is 1.10 g/cc. Although this is likely true for adults between the ages of about 20 and 50 years, the two-component model has serious limitations when measuring the body composition of elderly, children, and ethnic groups (Lohman, 1992). Variation in total body water and mineral content of these extreme groups varies from the values of the 20 to 50-year-old subjects and this affects the density of fat-free weight.

During childhood and the elderly years, the body is changing more dramatically. Changes in body water and bone mineral content alter the density of the fat-free component. As these values increase over reference values, there is a linear increase in percent body fat errors obtained with the two-component method. Lohman (1992) provides an excellent discussion on the effect of bone mineral differences on the accuracy of percent body fat determinations. The multicomponent model is a method of minimizing these errors. The multicomponent model not only includes body density but also water (w) and mineral (m) content. A multicomponent equation (Lohman, 1992) that can be used for children or adults of any age and any ethnicity is as follows—

#### Multicomponent Fat Model

(16)

$$\%Fat = \frac{2,749}{BD} - (0.727 \times w) - (1.146 \times m) - 2.053$$

As the percent body fat equation shows, body density is a primary element for measuring percent body fat with either the two-component or multicomponent models. The accurate measurement of percent fat depends on the accurate measurement of body density, and the underwater weighing method is the “gold standard.” Although many believe that measuring underwater weight is the biggest source of inaccuracy, this is not the case. The air component (i.e., residual lung volume and gastrointestinal air) is the most error-prone variable. Table 4.3 provides the potential problems associated with realistic measurement errors of the variables used to measure underwater percent body fat. Provided is the degree that actual percent body fat would vary for three different error conditions. Typically, the measurement error of underwater weight is less than 0.1 kg. Body weight and water temperature can be measured very accurately. Residual lung volume can be difficult to measure and the air in the gastrointestinal tract is estimated at 100 ml (0.1 L). Air component errors of  $\pm 0.1$  L translate to percent body errors of  $\pm 0.7\%$  fat, but air component errors of  $\pm 1$  L lead to huge percent body fat errors,  $\pm 8.0\%$  fat. Estimating residual volume from age, height, and gender yields air component errors in this magnitude (Morrow, Jackson, Bradley, & Hartung, 1986). If the underwater weighing method is to be used, the air component must be measured accurately.

**Height and Weight Assessment of Body Composition—**Because of the need for highly trained technicians and expensive laboratory equipment, hydrostatically determined body composition is rarely used in field settings. A common method is to use a weight-height ratio. Body mass index

**Table 4.3 Effects of component errors on underwater determined by Siri percent body fat"**

Underwater Weight Variable	Actual Value	Error Conditions		
		1	2	3
Air Component (L)	1.20	1.30	1.60	2.20
% Fat	15.00%	14.30%	12.20%	8.00%
Underwater Weight (kg)	3.36	3.38	3.41	3.46
% Fat	15.00%	14.90%	14.60%	14.30%
Body Weight (kg)	70.00	70.10	70.50	71.00
% Fat	15.00%	15.10%	15.30%	15.50%
Water Temp (C')	36.00	36.10	36.50	37.00
% Fat	15.00%	15.10%	15.10%	15.20%

\* Constructed from published data of Going 1996; Pollock and Wilmore 1984. From Baumgartner and Jackson (1999)

(BMI) is the weight-height ratio often used for large-scale public health studies. BMI (Equation 17) is computed from weight in kilograms and height in meters.

$$\text{Body Mass Index (BMI)} \quad (17)$$

$$BMI = \frac{\text{Weight}}{\text{Height} \times \text{Height}}$$

The BMI standards used to define overweight for the Healthy People 2000 Public Health program are 27.8 for men and 27.2 for women (USDHHS, 1990). The World Health Organization (WHO) uses BMI to evaluate degree of obesity (Bouchard & Blair, 1999). The WHO standards are—

- \* BMI 25.0 to 29.9 — Overweight (pre-obesity)
- BMI 30.0 to 34.9 — Obesity Class I
- BMI 35.0 to 39.9 — Obesity Class II
- BMI ≥ 40.0 — Obesity Class III

Another method of interpreting BMI is by its relationship with hydrostatically measured percent body fat. Medical researchers (Gallagher, 1996) have published a generalized equation for estimating percent body fat from BMI. They studied more than 700 African-American and white men and women who ranged in age from 20 to 94 years. The multicomponent model was used to measure percent body fat. They found that age, BMI, and gender (Female = 0, Male = 1) were significantly related to percent fat, but ethnicity was not. The equation for estimating percent body fat from these variables is as follows—

$$\text{Gallagher et al. Equation } (R = 0.819, SEE = 5.68\% \text{ fat}) \quad (18)$$

$$\%fat = (1.45 \times BMI) + (0.12 \times Age) - (11.61 \times Gender) - 10.02$$

**Skinfold Equations**—Several researchers have published regression equations with functions to predict hydrostatically measured body density from various combinations of skinfold measurements. Early researchers developed equations for homogeneous populations. These were termed

“population-specific” equations. More than 100 population-specific equations appear in the literature. The second trend was to use what is termed “generalized equations,” which are equations that can be validly used with heterogeneous samples. Population-specific equations were developed on small, homogeneous samples, and their application is limited to that sample.

The generalized equations were developed on large heterogeneous samples using models that accounted for the nonlinear relationship between skinfold fat and body density. Age was an important variable for generalized equations (Durnin & Rahaman, 1967; Jackson & Pollock, 1978; Jackson et al., 1980). The main advantage of the generalized approach is that one equation replaces several without a loss in prediction accuracy. A detailed discussion of population-specific and generalized equations can be found in other sources (Cureton, 1984; Jackson, 1984; Lohman, 1982).

A factor analysis (Jackson & Pollock, 1976) of skinfold and body circumference variables showed that skinfolds measured the same general body composition construct. This suggested that the sum of several skinfolds provided the most accurate estimate of the body fat construct. It was discovered that equations that use the sum of three skinfolds were highly correlated ( $r \geq 0.97$ ) with the sum of seven skinfolds (Jackson & Pollock, 1978; Jackson et al., 1980). This showed that the sum of three skinfolds could be used without the loss of accuracy, and the sum of three equations has become standard. Multiple regression models were used to develop generalized skinfold equations for men (Jackson & Pollock, 1978) and women (Jackson et al., 1980) using the sum of seven and three skinfold sites. The Jackson-Pollock databases were also used to develop the prediction equations used for the YMCA adult fitness test (Golding et al., 1989). Equations 19 and 20 show the Jackson-Pollock sum of three skinfold equations —

$$\begin{aligned} \text{Females: Triceps, Suprailium and Thighs ( } R = 0.84, \text{ SEE } = 0.009) \quad (19) \\ BD = 1.099421 - (0.0009928 \times C3) + (0.00000023 \times \Sigma 3^2) - (0.0001382 \times \text{Age}) \end{aligned}$$

$$\begin{aligned} \text{Males: Chest, Abdomen and Thigh ( } R = 0.91, \text{ SEE } = 0.008) \quad (20) \\ BD = 1.10938 - (0.0008267 \times C3) + (0.0000016 \times \Sigma 3^2) - (0.0002574 \times \text{Age}) \end{aligned}$$

These equations have been used extensively to estimate the body composition of men and women. The equations provide accurate models for estimating body density. The limitation of the generalized equations is for estimating percent body fat. Using the two-component percent fat equations of Siri, Brozek, and Schutte will only provide accurate percent body fat estimates when the subject's fat-free weight averages a density of 1.10 g/cc. This limits their use to subjects between the ages of 20 and about 50 years.

**Body Circumferences Prediction Models** — Body circumferences have also been used to assess body composition. The research method used was to estimate body density from combinations of body circumference measurements. Body circumferences are correlated with hydrostatically determined body density. Tran and associates (Tran & Weltman, 1989; Tran, Weltman, & Seip, 1988) published generalized equations for estimating hydrostatically determined body density from various combinations of circumference measurements. In 1981, the U.S. Navy changed from using height and weight standards to percent body fat estimated from body Circumferences (Hodgdon & Beckett, 1984a; Hodgdon & Beckett, 1984b). The variables used for the U.S. Navy equations are height,



abdomen circumference, hip circumference, and neck circumference. The U.S. Army also uses body circumference to estimate percent body fat. Table 4.4 gives the Military body composition equations.

**Table 4.4 U.S. Military body composition equations**

Army (Vogel et al., 1988)

Men  $R = 0.82$ ,  $SEE = 4.02$

$Percent\ fat = 76.5 \times Log_{10}(abdomen\ II^* - neck) - 68.7 \times Log_{10}(height) + 46.9$

Women  $R = 0.82$ ,  $SEE = 3.60$

$Percent\ fat = 105.3 \times Log_{10}weight - 0.200 \times wrist - 0.533 \times neck - 1.574 \times forearm + 0.173 \times hip - 0.515 \times height - 35.6$

Navy (Hodgdon and Beckett, 1984a, b) and Air Force

Men  $R = 0.90$ ,  $SEE = 3.52$

$Density = -0.191 \times Log_{10}(abdomen\ II - neck) + 0.155 \times Log_{10}(height) + 1.032$

$Percent\ fat = 100 \times [( \text{---} ) - 4.51]$

Women  $R = 0.85$ ,  $SEE = 3.72$

$Density = -0.350 \times Log_{10}(abdomen\ I \dagger - neck) + 0.155 \times Log_{10}(height) + 1.032$

$Percent\ fat = 100 \times [( \text{---} ) - 4.51]$

Marine Corps (Wright et al., 1980, 1981)

Men  $R = 0.81$ ,  $SEE = 3.67$

$Percent\ fat = 0.740 \times abdomen\ II - 1.249 \times neck + 40.985$

Women  $R = 0.73$ ,  $SEE = 4.11$

$Percent\ fat = 1.051 \times biceps - 1.522 \times forearm - 0.879 \times neck + 0.326 \times abdomen\ II + 0.597 \times thigh + 0.707$

NOTE: Circumference measurements and height are in centimeters. SEE, standard error of the estimate.

\* Abdomen II is the circumference, measured in transverse plane, at the level of the umbilicus.

† Abdomen I is the "natural waist" and is defined as the smallest circumference, measured in the transverse plane, obtained between the lower margin of the xiphoid process and the umbilicus.

SOURCE: Adapted from Hodgdon (1992)

**Waist-Hip Ratio**—Medical research has shown that people with central, visceral types of obesity are particularly at risk for developing cardiovascular disease, stroke, and noninsulin-dependent diabetes mellitus. The field test used to measure central visceral obesity is waist-hip ratio (WHR) (Equation 21). Efforts are now underway to obtain better estimates of central visceral fat with imaging methods such as computed axial tomography (CT scans) (Wilmore et al., 1999). The development of central, visceral obesity is believed to be caused by an alteration in the body's metabolic system. Several of these endocrine abnormalities are associated with insulin resistance that is believed to be the cause of the increased disease risk. Although the BMI has been used to identify overweight individuals, Bray (Bray, 1993) proposed that both BMI and WHR be used to define health risk.

$$WHR = \frac{Waist-C}{Hip-C}$$

**Dual Energy X-Ray Absorptiometry** — Dual energy x-ray absorptiometry (DXA) is a method that evolved from the widespread use of single- and dual-photon absorptiometry. The development of computer technology enhanced the application of DXA technology to the measurement of body composition. Lohman (Lohman, 1996) reports that DXA can be used to measure total body and regional body composition, including the estimation of bone mineral content, lean tissue mass, fat-free mass and fat mass. Many believe that DXA may become a reference method of estimating human body composition and even replace underwater weighing. The DXA standard errors typically range from 2.5 to 3.5 percent when estimating percent fat from body density using the multicomponent method (Lohman, 1996). With improved computer technology, test methods and lower costs, DXA will likely grow in popularity. A major limitation of the hydrostatic method is that it can be difficult and even impossible to underwater-weigh persons who have a fear of water.

**Bioelectrical Impedance Method** — Bioelectrical impedance analysis (BIA) is based on the principle that the electrical resistance of the body to a mild electric current is related to total body water. Total body water and fat-free weight are highly related. The BIA method is simple and requires only the placement of four electrodes, two on the subject's ankle and two on the wrist. An electrical current is transmitted into the subject, and the resistance in ohms is read directly into a microcomputer that calculates body composition.

In the early stages of BIA technology, the accuracy of BIA was a major concern. One study showed that this method was no more accurate than BMI (Jackson, Pollock, Graves, & Mahar, 1988). Recent research (Lohman, 1992) showed that with suitable equations, BIA estimates of percent body fat are similar in accuracy to skinfold estimates, except for the obese and very lean. Equations developed on the general population tend to underestimate percent body fat of the obese and overestimate the percent body fat of very lean subjects, showing that more research is needed to develop generalized BIA equations.

**Percent Body Fat Standards** — Different percent body-fat standards are needed for men and women. Women not only have a higher percentage of their weight in storage fat measured with the caliper but also in essential fat consisting of lipids of the bone marrow, central nervous system, mammary glands, and other organs. Because of this additional storage fat, the percent body fat of women tends to be about 7 to 8 percent higher than that of men's (Jackson & Pollock, 1978; Jackson et al., 1980). Tables 4.5 and 4.6 give the percent body fat standards recommended by the American College of Sports Medicine (Gettman, 1993).

## Muscular Strength

Muscular strength is the maximum amount of force that a muscle group can exert. Muscle contractions can be either dynamic or static. Static contractions do not involve movement and are called isometric. Dynamic contractions involve movement, either concentric, the muscle shortens,

**Table 4.5 Percent body fat standard for women contrasted by age group'**

Standard	Age Group (Women)			
	≤30	30–39	40–49	≥50
High	>32%	>33%	>34%	>35%
Moderately High	26–32	27–33	28–34	29–35
Optimal Range	15–25	16–26	17–27	18–28
Low	12–14	13–15	14–16	15–17
Very Low	≤11%	212%	113%	214%

\* Reprinted, by permission, from L. R. Gettman, 1993, American College of Sports Medicine: Resource manual for guidelines for exercise testing and prescription (2nd ed. pp. 229–246). Philadelphia: Lea & Febiger.

**Table 4.6 Percent body fat standards for men contrasted by age group\***

Standard	Age Group (Men)			
	≤30	30–39	40–49	≥50
High	>28%	>29%	>30%	>31%
Moderately High	22–28	23–29	24–30	25–31
Optimal Range	11–21	12–22	13–23	14–24
Low	6–10	7–11	8–12	9–13
Very Low	≤5%	26%	27%	≤8%

or eccentric, the muscle lengthens. The dynamic forms include isotonic and isokinetic. Isotonic involves moving a weight against gravity. Lifting the weight uses a concentric contraction while lowering the weight uses an eccentric contraction. Isokinetic involves muscle contractions at a fixed speed. Strength testing may be either open or closed kinetic chain. An open kinetic chain is when the end of the limb segment is free in space while a closed kinetic chain is when the end segment or joint meet with external resistance that prevents or restrains free motion. In a closed kinetic chain, movement at one joint produces movement at the other joints in the chain or system. In an open kinetic setting, the distal limb segment can move freely (Baumgartner & Jackson, 1999). Table 4.7 provides an overview of the strengths and weaknesses of the strength testing method (Baumgartner & Jackson, 1999).

**Types of Strength Tests** — Isometric strength testing has historically been popular. It is the maximum force that a muscle group can exert without movement. Mechanical devices such as tensiometers and spring dynamometers were the units first used to measure the force applied during an isometric contraction. These mechanical units have been replaced with electronic load cells. Professional standards for equipment and isometric test methods are published (Chaffin, 1975; NIOSH, 1977). Isometric tests can measure specific muscle groups or combinations of muscle groups. Isometric arm, shoulder, torso, and leg strength have been used extensively for preemployment decisions (Baumgartner & Jackson, 1999; Jackson, 1994; NIOSH, 1977).

**Table 4.7 A comparison of strength testing methods\***

Method	Strengths	Weaknesses
isometric	<ol style="list-style-type: none"> <li>1. Moderately inexpensive</li> <li>2. Can be used to test a variety of different muscle groups</li> <li>3. Closed kinetic chain</li> <li>4. Strong research base for preemployment testing</li> <li>5. Normative data are available</li> <li>6. Takes very little time to test a subject</li> <li>7. Easy to learn how to administer the tests</li> </ol>	<ol style="list-style-type: none"> <li>1. Only one joint angle is tested</li> <li>2. Does not provide a torque strength curve</li> <li>3. Cannot measure dynamic contractions</li> </ol>
Isotonic	<ol style="list-style-type: none"> <li>1. Very inexpensive</li> <li>2. Many different types of equipment can be used</li> <li>3. Closed kinetic chain</li> <li>4. Tests often duplicate strength development program</li> <li>5. Takes very little time to test a subject</li> <li>6. Easy to learn how to administer the tests</li> </ol>	<ol style="list-style-type: none"> <li>1. Never measure "true maximum"</li> <li>2. Cannot obtain a strength curve</li> <li>3. Risk of injury if free weights are used</li> <li>4. Different types of equipment affect the score, need equipment specific norms</li> <li>5. Can be difficult to find 1-RM</li> </ol>
Isokinetic	<ol style="list-style-type: none"> <li>1. Can obtain strength curves for many different speeds</li> <li>2. Can obtain both eccentric and concentric contractions</li> <li>3. Data can be expressed in many different ways</li> <li>4. Valuable for rehabilitation process</li> </ol>	<ol style="list-style-type: none"> <li>1. Very expensive equipment</li> <li>2. Not closed kinetic chain</li> <li>3. Velocity of movement affects torque output, need norms for various speeds</li> </ol>

\* From Baumgartner and Jackson (1999)

Isotonic strength is measured by determining the maximal force that a muscle group can exert with a single contraction. An isotonic strength test measures the maximum weight that can be lifted with a single repetition. This is the one-repetition maximum test (1-RM). The equipment used to measure 1-RM strength includes free weights and progressive resistance equipment. The most difficult part of the test is to find the subject's maximal load. Several different weights will need to be tried to find the proper 1-RM weight. Bench and leg press 1-RM strength are common isotonic tests (Gettman, 1993).

Isokinetic methods measure torque through a defined range-of-motion while keeping the speed of movement constant. The equipment used to measure isokinetic strength is a load cell interfaced with a computer. The computer unit controls the speed of movement and measures torque. This yields the muscle group's torque curve for the selected constant velocity. Both muscle strength and the velocity of movement affect the shape and magnitude of the curve. As the muscle contracts at a faster rate, it cannot generate as much torque so a lower curve is obtained. Test results from different test centers are not comparable unless the sites used the same equipment and the same test velocity.

Isokinetic equipment is expensive and so is usually used only at well-equipped testing centers such as sports medicine and physical therapy facilities. Over the past few years isokinetic testing has lost favor. There are several reasons for this. Major factors are the cost of the equipment and changes in our health care systems. Managed health care corporations have dramatically reduced the money they will pay for strength evaluations. Another reason is that isokinetic tests are largely open-kinetic chain, and the current rehabilitation philosophy is to use closed-kinetic chain. Isokinetic tests are typically used to measure isolated muscle groups like knee extension and flexion.

**Correlations Among Types of Strength Tests** — Although strength tests involve dynamic and static contractions, they are highly correlated. In a controlled laboratory study, Laughlin (1998) examined the relationship between closed-kinetic isometric and dynamic 1-RM leg strength. A sample of 57 healthy female athletes was administered isometric and isotonic leg strength tests. A Cybex

leg press machine was used to test the athletes' maximum isotonic leg strength. Both the dominant and nondominant legs were tested. The Cybex unit was also used to measure dominant and non-dominant isometric strength at 30°, 60°, and 90° knee flexion. Table 4.8 gives the results of the study. The correlations between the isometric and isotonic strength tests were very high. The coefficients ranged from 0.91 to 0.94. These high correlations were expected because the test positions were duplicated, thereby increasing the probability that the isometric and isotonic tests measured the strength of the same muscle groups. These data showed that the type of muscle contraction, static and dynamic, did not affect the measurement of strength.

**Table 4.8 The Pearson product-moment Correlations between static and dynamic leg strength**

Isometric Test	Isotonic Test	
	Dominant Leg	Non-Dominant Leg
90° Knee Flexion	.94	.91
60° Knee Flexion	.93	.91
30° Knee Flexion	.93	.94

## Muscular Endurance

Muscular endurance is the ability to persist in physical activity or to resist muscular fatigue. Endurance tests can measure absolute endurance where the power output is the same for all subjects tested, or relative endurance where the power output varies among the subjects tested. Absolute endurance tests tend to be correlated with strength while the correlation between relative endurance and strength tests tend to be close to zero (Baumgartner & Jackson, 1999; deVries & Haush, 1994; Jackson, Osburn, & Laughery, 1984; Jackson, Osburn, Laughery, & Vaubel, 1992; Jackson, Osburn, & Laughery, 1991a; Jackson, Osburn, Laughery Sr., & Vaubel, 1991b). There are two general types of endurance tests: laboratory tests that use an ergometer (cycle or arm) to stan-

**Table 4.9 Factor analysis of strength data published by Dempsey et al., 1998**

Test	Type of Strength	Factor Loading
Isometric 15'	Static	.70
Isometric 75*	Static	.83
Isokinetic 0.1**	Dynamic	.92
Isokinetic 0.2**	Dynamic	.90
Isokinetic 0.4**	Dynamic	.93
Isokinetic 0.6**	Dynamic	.89
Isokinetic 0.8**	Dynamic	.90
Incrementally Lifting	Dynamic	.70
Power	Dynamic	.87
MAWL	Psychophysical	.83

\* Height of lift position.

\*\* Speed of movement

standardize power output; and field tests such as push-ups, pull-ups, or sit-ups that are common items of fitness test batteries.

**Ergometer Tests**—An all-out cycling endurance test first described in 1973 was called the Katch test (McArdle, Katch, & Katch, 1991). This test, refined at the Department of Research and Sport Medicine at the Wingate (Israel) Institute, is now known as the Wingate anaerobic power test (Bar-Or, 1987). This has become the test of choice for measuring anaerobic endurance. The Wingate power test involves cycling as fast as possible for 30 seconds at a set resistance, which is a proportion of the subject's body weight (Bar-Or, 1987; Baumgartner & Jackson, 1999). Arm endurance can be measured in the same way with an arm ergometer. Arm ergometer tests have been used in employment settings (Laughery & Jackson, 1985).

**Field Tests**—Push-ups, pull-ups, flexed-arm hang, and sit-ups are common muscular endurance field tests. The tests of this ability require the subject to move or support their body weight against the pull of gravity. This may involve either isometric or isotonic contractions. The tests are either performed to exhaustion (e.g., number of pull-ups completed) or for a specified duration of time (e.g., number of sit-ups completed in two minutes). There is a negative correlation between body weight and this basic physical ability. The correlation tends to be higher between percent of body fat and this ability (Baumgartner & Jackson, 1999).

## Flexibility

Flexibility is the range of movement about a joint. Individual differences in flexibility depend on physiological characteristics that influence the extensibility of the muscles and ligaments surrounding a joint. Although it is agreed that certain levels and types of flexibility are desirable, the

degree of flexibility desired is yet to be determined. Typically, a trunk flexibility is included as an item in fitness test batteries (Baumgartner & Jackson, 1999; Golding et al., 1989).

Flexibility is often regarded as a single general factor or ability. Harris (Harris, 1969) conducted a factor analysis study to determine whether flexibility is a single or general factor. Two types of flexibility tests were used in this study—

1. tests that measure the movement of a limb involving only one joint action, and
2. composite measures of movements that require more than one joint or more than one type of action within a single joint.

The analysis revealed many intercorrelations to be near zero, which implies specificity instead of generality. A factor analysis of these data revealed 13 different factors of flexibility. Harris concluded, then, that there is no evidence that flexibility is a single general factor. These data suggest that flexibility tests should duplicate the specific type of flexibility desired.

---

## Work Sample Tests

This section reviews the advantages and disadvantages of work sample tests. A work sample test is designed to duplicate or simulate a critical work task or a series of important work tasks. The advantage of a work-sample test is that it simulates the actual working conditions. Although criterion-related and construct validation methods can be used to examine the validity of a work sample test, the primary method used is content validity. Work sample tests are commonly used to screen applicants for police officer and firefighter jobs. Arvey (Arvey, Nutting, & Landon, 1992) reported that most police and firefighter physical ability tests consist of some combination of job sample tests. Examples of common firefighter work sample test items (Davis et al., 1992) are as follows —

- \* Stair Climb — Carry a 58-pound hose bundle up 5 flights of stairs.
- **Hoseline Drag** — Drag a 1.75-inch charged hose 100 feet.
- **Rescue Dummy Drag** — Lift and drag a 175-pound dummy 100 feet.
- **Smoke Extractor Carry** — Lift and transport a 47.5-pound fan a distance of 150 feet.
- **Kieser Force Machine** — Repeat pounding an object with a sledge hammer until it is moved a specified distance.

This last test simulates a forced entry into a building.

The major advantage of a good work sample test is that it has the potential of enhancing its content validity by duplicating the actual work task. To illustrate, assume that a job requirement is to lift a 75-pound box from floor-to-knuckle height. An example of a content valid work sample test would be one that requires the subject to lift a 75-pound box from the floor and place it on a table. The test would be objectively scored as a pass or fail.

Although work sample tests have excellent content validity, they also have limitations. In some instances, work sample tests are expensive to create and difficult to set up and transport from one test location to another site. For instance, a wooded structure was made to duplicate the physical environment of the cargo space of an aircraft used to transport freight (Jackson, Osburn, Laughery, & Young, 1993). Along with the test administration difficulties, Ayoub (Ayoub, 1982) maintains that work sample tests have two major limitations. The first limitation is safety. Applicants seeking employment are likely to be motivated to pass the work sample test. Highly motivated applicants who lack the physical capacity to perform the test are likely to increase their risk of injury. Research shows that the risk of injury increases as the workers perform materials handling tasks that approach their maximum capacity (Dehlin, Hendenrud, & Horal, 1976; Herrin et al., 1986; Magora, 1970; Snook, Campanelli, & Hart, 1978). As another example, outdoor telephone craft jobs require employees to climb telephone poles, and accident data showed that this was a dangerous task (Reilly, Zedeck, & Tenopir, 1979). Using a pole-climbing test to screen applicants would have content validity, but would likely be dangerous for untrained employees.

A second limitation of work sample tests is that they do not give any information about the applicant's maximum work capacity (Ayoub, 1982). A work sample test is often scored by pass or fail (e.g., lifted a 95-pound jackhammer and carried it a specified distance). Some applicants can easily complete the test, while others may just pass and be working near their maximum. If it can be assumed that there is a linear relationship between job performance and the preemployment test performance, applicants with the highest test scores can be expected to be the more productive workers. Testing for maximum capacity not only identifies the most potentially productive workers but also provides the opportunity to define a level of reserve needed to reduce the risk of musculoskeletal injury.

Another potential limitation of work sample tests is that there can be a factor of skill required by the test. This can be illustrated with an example. The ability to use a sledge hammer is a work sample test item used to select firefighters. Sledge hammering is a physically demanding task in which the ability to deliver a forceful blow not only depends on strength but also on neuromotor coordination of the stroke. If a person has never used a sledge hammer, he or she would not have yet developed a coordinated stroke. Part of firefighter training is to develop the skill that is used to gain forced entry.

---

## Relationship of Basic Abilities to Work Sample Tests

Although work sample tests can have the advantage of being content valid, they can be difficult to set up and time consuming to administer. To overcome this limitation, researchers have examined the relationship between the basic ability tests and the work sample tests. When the basic ability tests and the work sample tests were highly correlated, the basic ability tests replace the work sample tests. Arnold and associates (Arnold, Rauschenberger, Soubel, & Guion, 1982) were among the first to demonstrate that basic ability tests were highly correlated with work sample tests. They developed work sample tests for steelworkers and administered them to 81 women and 168 men who were in their first six months of employment at three different plant locations. In addition, the worker's strength, flexibility, agility, balance, and cardiorespiratory endurance were tested. They discovered that arm dynamometer strength was highly correlated with work sample test performance.



The zero-order correlations between arm strength and work sample test performance were consistently high, 0.82, 0.85, and 0.85 for the three sites. Multiple regression analysis showed that flexibility, agility, balance, and cardiorespiratory endurance did not account for an additional significant proportion of work test variance. Although just strength was correlated with the steelworker work sample tests, it is important to realize this may not be true for other work sample tests. There is considerable evidence that  $\text{VO}_2\text{max}$ , strength, and fat-free weight are significantly related with work sample test performance. Provided next is a brief review of the role of  $\text{VO}_2\text{max}$ , strength and fat-free mass on work sample test performance.

## **Role of $\text{VO}_2\text{max}$**

Work tasks such as climbing stairs and fighting fires have a significant aerobic endurance component. Published physiological research documents the cardiovascular response of these work tasks. A tradition of work physiology has been to define the energy cost of work tasks with oxygen uptake (Durnin & Passmore, 1967; McArdle et al., 1991; Wilmore & Costill, 1994). This provides a sound physiological basis for establishing a cut-score.

A current, important research focus is to define the energy cost needed to fight fires. This research focus can be attributed to litigation leveled at the validity of firefighter preemployment tests and the use of age to terminate employment. Several investigators (Barnard & Duncan, 1975; Lemon & Hermiston, 1977a; Lemon & Hermiston, 1977b; Manning & Griggs, 1983; O'Connell, Thomas, Caddy, & Karwasky, 1986; Sothmann, Saupe, Jasenor, & Blaney, 1992) published data showing that fire-suppression work tasks have a substantial cardiovascular endurance component. In an important study, Sothmann and a team of researchers (Sothmann et al., 1990) provide strong evidence that the minimum  $\text{VO}_2\text{max}$  required to meet the demands of fire fighting is 33.5 ml/kg/min. The authors used a work sample test involving seven job-related firefighter tasks. The sensitivity (percentage of correctly classified unsuccessful performers) and specificity (percentage of correctly classified successful performers) for a  $\text{VO}_2\text{max}$  cut-score of 33.5 ml/kg/min was 67 percent and 83 percent. Lowering the cut-score to 30.5 ml/kg/min dropped the sensitivity to 25 percent and increased the specificity to 95 percent.

Because of the expense involved, laboratory  $\text{VO}_2\text{max}$  tests are rarely used to make employment decisions. Rather, the 1.5-mile distance run tends to be used to evaluate aerobic capacity in field settings. In a recent firefighter preemployment test (Jeanneret & Associates, 1999), the correlation between 1.5-mile run time and elapsed time to complete a five-item firefighter work sample test was 0.59 ( $n = 125$ ), supporting the laboratory data showing the fire-fighting ability was a function of aerobic power.

## **Role of Strength and Fat-Free Weight**

Strength is the basic ability most often shown to be associated with physically demanding work tasks (Hogan, 1991). Since the body composition component of fat-free weight consists mainly of muscle mass, the body's force-producing component, the correlation between strength and fat-free

mass is high. It has been shown that the correlation between isometric strength and absolute fat-free weight exceeds 0.80 and both are equally valid predictors of lifting capacity (Jackson, Borg, Zhang, Laughery, & Chen, 1997).

**Role of Strength on Work Sample Tests**—Researchers' from the University of Houston and Rice University, Houston, Texas, completed a series of preemployment studies in which one step in the research methodology was to examine the relationship between isometric strength and work sample test performance. Although the complete research methods appear in technical reports, a representative sample of this research has been published (Jackson et al., 1999; Jackson, Borg, Zhang, Laughery, & Chen, 1996a; Jackson et al., 1997; Jackson et al., 1984; Jackson et al., 1992; Jackson et al., 1991a). The results of this research are consistent with that reported by Arnold and Associates (Arnold et al., 1982) showing that muscular strength was an important determinant of content-valid work sample tests.

Table 4.10 gives the product-moment correlation between isometric strength tests and work sample test performance. The strength tests were highly correlated with several different types of work sample tests. The highest correlations were between isometric strength tests and static strength work sample tests that involved the capacity to generate force in various body positions. These work sample tests measured the subject's capacity to generate push, pull, and lift force and generate valve cracking force while in common positions required of refinery workers. The isometric strength tests were also correlated with work sample absolute endurance tests. The dynamic tasks were either performed to exhaustion (e.g., valve turning) or at a "comfortable rate" set by the person being tested (e.g., shoveling 600 pounds of material over a 3.5-foot wall). These dynamic work sample tests involved valve turning, shoveling, and repetitive materials handling tasks. The correlations between strength and the absolute endurance work tests ranged from 0.67 to 0.83. Besides these work tasks, strength has been shown to be related to lifting capacity (Jackson et al., 1999; Jackson et al., 1997) and firefighter work sample test performance (Jeanneret & Associates, 1999). The correlations between the elapsed time to complete the five-item firefighter work sample test, and four isometric strength tests ranged from -0.61 for leg and torso strength to -0.90 for arm strength.

**Role of Body Composition on Work Sample Tests**—Percent body fat and fat-free mass are the body composition variables most often used to examine their relationship to work sample test. Percent body fat tends to be negatively correlated with work sample tests that require the subjects to move their body mass. Since fat-free mass is the body's force-producing component, it tends to correlate with work tasks related to strength. Skinfold estimated percent body fat was found to be significantly related to the ability of telephone workers to climb poles (Bernauer & Bonanno, 1975; Reilly et al., 1979). Triceps skinfold thickness was one of the tests used to screen applicants for pole-climbing school. A limitation of using skinfold fat or percent body fat in employment decisions in the public sector is that the chance of litigation increases. Women have about 5 to 10 percent more body fat than men, which is largely because of differences in essential fat (Lohman, 1992). Using a single skinfold site such as triceps also invites litigation because of male and female differences in fat patterning (Jackson & Pollock, 1978; Jackson et al., 1980).

Hodgdon and Associates (Hodgdon, 1992) examined the relationship between body composition, fitness, and materials handling tasks required of U.S. Navy enlisted men. The two work sam-

**Table 4.10 Correlations between the Sum of Isometric Strength\* and Simulated Work Sample Tests**

Reference Work	Sample Test	Type of Test	$r_{xy}$
Jackson and Osburn 1983	One-arm Push Force	isokinetic – Peak Torque	0.91
Jackson 1986	Push Force	Static – Max Force	0.86
Jackson et al., 1993	Push Force	Static – Max Force	0.78
Jackson 1986	Pull Force	Static – Max Force	0.78
Jackson et al., 1993	Pull Force	Static – Max Force	0.67
Laughery and Jackson 1984	Lifting Force	Static – Max Force	0.93
Jackson et al., 1998	valve Cracking	Static – Max Force	0.91
Jackson et al., 1992	Valve-Turning	Dynamic – Endurance	0.83
Jackson et al., 1993	Box Transport	Dynamic – Endurance	0.76
Jackson et al., 1992	Moving Document Bags	Dynamic – Endurance	0.70
Jackson et al., 1991	Shoveling Coal	Dynamic – Endurance	0.71
Jackson et al., 1991	50-pound Bag Carry	Dynamic – Endurance	0.63
Jackson and Osburn 1983	70-pound Bag Carry	Dynamic – Endurance	0.87

From Baumgartner and Jackson (1999)

ple materials handling tasks were the maximum weight of a box that could be lifted to elbow height, and the total distance a 34-kilogram box could be carried during two 5-minute work bouts. The variable most highly correlated with maximum box lift was fat-free mass, 0.84. The variables most highly correlated with the box-carry test were push ups, 0.56; 1.5-mile run time, -0.67; and fat-free mass, 0.44. Fat-free mass was highly correlated with muscular strength measures and suggested the possibility of using fat-free mass as an approximation of general strength in job assignment.

Vogel and Friedl (Vogel & Friedl, 1992) examined the relationship between body composition and absolute lifting capacity. They reported significant correlations between maximum lifting capacity and fat-free mass for male and female soldiers. The slope of the male equation was 2.2 times larger than the slope of the women's equation, suggesting the lack of homogeneity of male and female regression lines. Other research (Jackson et al., 1997) also showed that the relationship between fat-free weight and lift capacity was a function of muscular strength.

## Need for Further Research

Published data show that there tends to be a significant relationship between various basic ability tests and work sample tests. These correlations range from low to high, depending on the tests studied. A limitation of this body of research is that many basic ability tests tend to be correlated with each other, making it difficult to define the true determinants of work test performance. As one example, percent body fat tends to be significantly correlated with tests involving movement of the body. This includes not only running tests of speed and aerobic endurance, but also arm tests such as chin-ups and push ups that involve repeatedly moving the body mass with the arms.

A major problem that needs further study is the method used to scale  $\text{VO}_2\text{max}$ . Maximal oxygen uptake can be expressed in either absolute or relative terms. Absolute  $\text{VO}_2\text{max}$  is the total volume of oxygen the person can consume during exhausting work, and is expressed in a metric of milliliters of oxygen per minute (ml/min) or liters of oxygen used per minute (L/min). In contrast, relative  $\text{VO}_2\text{max}$  is expressed in a metric normalized for body weight, per kilogram of body weight (ml/kg/min). The metric used to express  $\text{VO}_2\text{max}$  affects test validity. Table 4.11 illustrates the problem and shows the differences when  $\text{VO}_2\text{max}$  is expressed in absolute and relative differences. The table provides the correlations between body composition variables and  $\text{VO}_2\text{max}$  expressed in relative and absolute terms. Data are provided for men and women from a very large database of NASA/JSC (Houston, TX) employees (Jackson et al., 1995; Jackson et al., 1996b).  $\text{VO}_2\text{max}$  was metabolically measured following standard procedures for defining  $\text{VO}_2\text{max}$ , and skinfold equations (Jackson & Pollock, 1978; Jackson et al., 1980) were used to measure the body composition variables. These data show that when  $\text{VO}_2\text{max}$  is expressed in relative terms (ml/kg/min), body weight, percent body fat, and fat weight are negatively correlated with aerobic fitness. In contrast, when expressed as absolute  $\text{VO}_2\text{max}$  (ml/min), fat-free weight is the variable most highly correlated with aerobic fitness.

**Table 4.11 Product-moment correlations between body composition variables and  $\text{VO}_2\text{max}$  expressed in relative (ml/kg/min) and absolute (ml/min) terms**

Variable	Males (n = 1,477)		Females (n = 409)	
	$\text{VO}_2\text{max}$ (ml/kg/min)	$\text{VO}_2\text{max}$ (ml/min)	$\text{VO}_2\text{max}$ (ml/kg/min)	$\text{VO}_2\text{max}$ (ml/min)
Weight	-0.30*	0.32'	-0.48'	0.14*
% Fat	-0.62'	-0.24'	-0.48''	0.14'
Fat-Free Weight	-0.04	0.57'	-0.05	0.51*
Fat Weight	-0.56'	-0.06	-0.67'	-0.16*

\*  $p < 0.01$

An issue that needs further research concerns the  $\text{VO}_2\text{max}$  metric that gives the highest validity of work test performance. The typical method used to express aerobic fitness is in relative terms (ml/kg/min). The data provided in Table 4.11 suggests that relative  $\text{VO}_2\text{max}$  would be the most valid when examining work tasks that involve repetitive bodily movement such as climbing towers, walking, and running long distances. The high correlation between percent body fat and relative  $\text{VO}_2\text{max}$  should be expected because transporting high levels of fat makes body movement tasks more demanding. It has been demonstrated that endurance work tasks of shoveling, repetitive material transports (Jackson et al., 1991a) and valve turning (Jackson et al., 1991b) were correlated with strength. This link can be traced to the correlation between fat-free weight and absolute  $\text{VO}_2\text{max}$ . The correlation between fat-free weight and strength is reported to exceed 0.80 (Jackson et al., 1997).

Relative  $\text{VO}_2\text{max}$  is the metric most often used to examine physically demanding tasks such as firefighting (Sothmann et al., 1990). Although firefighting tasks involve repetitive bodily movements, they also require transporting materials of a constant weight, such as a 58-pound hose bundle. Assuming the same percent body fat and relative  $\text{VO}_2\text{max}$ , transporting a constant weight load would be less demanding for someone with more fat-free mass than for someone with less fat-free

mass. For the same percent body fat and relative  $\text{VO}_2\text{max}$ , persons with greater fat-free mass have a higher absolute  $\text{VO}_2\text{max}$ . This suggests that absolute  $\text{VO}_2\text{max}$  (L/min or ml/min) would be a more valid test than relative  $\text{VO}_2\text{max}$  (ml/kg/min) to evaluate a worker's aerobic capacity to perform repetitive material handling tasks with a constant load for all workers.

Davis and Associates (Davis, 1992) obtained  $\text{VO}_2\text{max}$  data on a sample of 25 firefighters who also completed a five-item work sample firefighter test. An analysis these data showed that the correlation between the work sample test and relative  $\text{VO}_2\text{max}$  was -0.23 and not statistically significant ( $p = 0.281$ ). In contrast, the correlation between absolute  $\text{VO}_2\text{max}$  and work sample test performance was higher, -0.42, and statistically significant ( $p = 0.036$ ). This showed that absolute  $\text{VO}_2\text{max}$  was a more valid test than relative  $\text{VO}_2\text{max}$ , suggesting that the determinate firefighter work sample test performance depended on both aerobic endurance and muscle mass.

The data from a recent preemployment firefighter study (Jeanneret & Associates, 1999) provide additional insight into these complex relationships. A total of 125 firefighters (31 women and 94 men) completed the five-item work sample firefighter test used by Davis and associates (Davis, 1992). The firefighters' aerobic fitness was assessed with the 1.5-mile walk/run test, which is a field test of relative  $\text{VO}_2\text{max}$ . Besides the aerobic tests, the firefighters' percent fat was measured from skinfold fat (Jackson & Pollock, 1978; Jackson et al., 1980) and their strength evaluated with arm, shoulder, torso, and leg isometric strength tests (Baumgartner & Jackson, 1999). Table 4.12 provides the multiple regression analyses examining the role of relative  $\text{VO}_2\text{max}$  and combination of strength or fat-free weight on log transformed<sup>3</sup> firefighter work sample test performance. The unstandardized (b) and standardized ( $\beta$ ) regression equations are provided. One model used 1.5-mile combined with fat-free weight, and the other replaced fat-free weight with strength. The two regression models produced very similar results. The log transformed firefighter work sample test performance was an independent function of aerobic fitness and muscle mass. Fat-free mass and the sum of the four strength tests were two different measures of muscle mass sampled two ways, by fat-free mass and the sum of the four strength tests.

**Table 4.72 Multiple regression analysis comparing the effect of fat-free weight and strength in combination with 7.5-mile run and firefighter work sample test performance**

Variable	Fat-Free Model		Strength Model	
	b	$\beta$	b	$\beta$
intercept	0.816*		0.723'	
1.5-mile Run	0.053*	0.497'	0.043*	0.418'
Fat-Free Weight	-0.005*	-0.683*		
$\Sigma$ Four			-0.001*	-0.612'
R	0.91*		0.85*	

\*  $p < 0.01$

Many physically demanding jobs are repetitive materials handling tasks. Often, tasks are termed "endurance" or "strength" tasks. The data provided in this section illustrate that these tasks are likely to be more complex and a function of a combination of different basic abilities. Additional research is needed to define the role of aerobic fitness, body composition, and muscular strength

variables on work task performance. Not only would this provide a better understanding of the work task performance it would also provide the scientific value of using combinations of basic ability tests for selection and placement decisions.

---

## **Relationship of Basic Abilities on Health**

A suitable level of physical fitness is not only important for performing work tasks but it also is important for decreasing the risk for degenerative diseases and work-related musculoskeletal injury. This section provides a brief discussion on the role of aerobic fitness and body composition on health, and function of strength on musculoskeletal injury. This is covered in more detail in the health chapter of this report.

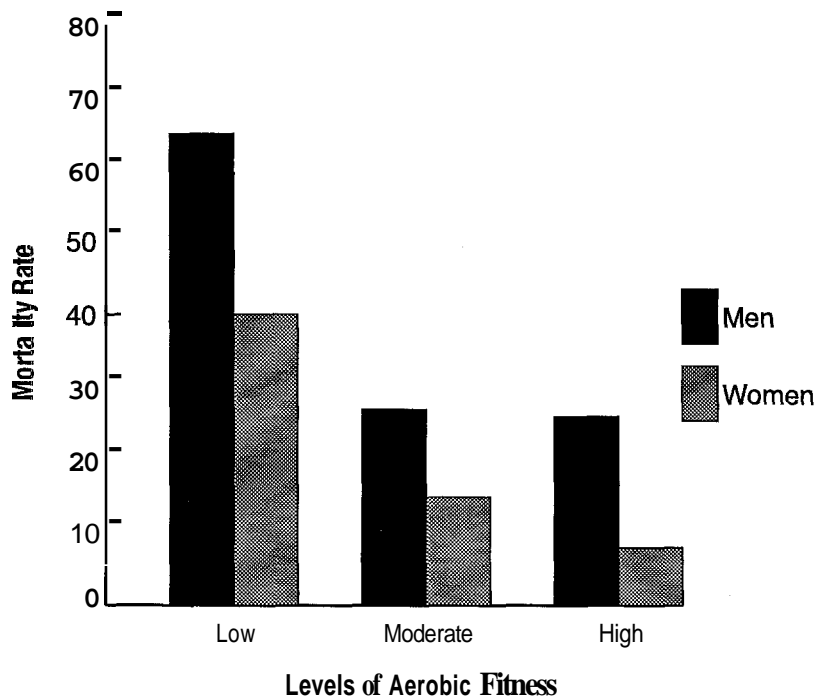
### **Aerobic Fitness and Health**

The classic study from the Institute for Aerobics Research (Dallas, TX) showed that low aerobic fitness predicted high mortality rates (Blair et al., 1989). The participants of the study were healthy men and women who were free from diseases such as high blood pressure or diabetes. After a maximal treadmill test, the participants were followed for several years. Figure 4.3 provides a graphic summary of the study. The greatest drop in death rate was between the lowest and moderate fitness groups. The death rates of the moderate and high fitness groups were more similar. This study shows the beneficial effect of a moderate level of aerobic fitness on mortality. The “low fit” group was the 20 percent of the men and women with the lowest age-adjusted aerobic fitness.

In a second study, the researchers discovered that changes in fitness were related to changes in mortality risk (Blair et al., 1995). Those who improved their aerobic fitness by moving from the low to the moderate or high categories reduced their future risk of death. Moderate fitness levels can be attained for most people who engage in regular aerobic exercise by doing the equivalent of walking about three miles a day. In a more recent study (Lee et al., 1999), low aerobic fitness was a more powerful risk factor of all-cause mortality than obesity. The Institute for Aerobics data provide solid epidemiological evidence showing that low aerobic fitness is a major risk factor of both cardiovascular diseases and all-cause mortality.

### **Body Composition and Health**

The relationship between weight and mortality is “J” shaped (Lew & Garfinkel, 1979). Being overweight is associated with many medical problems such as hypertension, diabetes, and heart disease. These illnesses lead to increased morbidity and reduced longevity. Overweight people are more likely to be diabetic, hypertensive, and have higher cholesterol levels. Since these are major, independent cardiovascular disease risk factors, some believe that the increased chance of cardiovascular disease associated with being overweight is due just to these other cardiovascular disease



**Figure 4.3** *The participants in the lowest aerobic fitness group had the highest death rates. The mortality rate of the moderate and highly fit group was about the same. The high mortality rate occurs mainly in the low fitness group. Graph made from published data (Blair et al., 1989). From Baumgartner and Jackson (1999).*

risk factors. This is not true. Being overweight puts you at a higher risk of heart disease and stroke (Hubert et al., 1983). Not only did the data from the Framingham study (Hubert et al., 1983) establish that obesity was a major, independent risk factor for cardiovascular disease, it also showed that gaining weight resulted in a higher risk while losing weight lowered health risk.

Although the relationship between obesity and cardiovascular disease is well documented, recent data suggest that obesity plays a role in breast cancer, the most common cancer of women. The causes of breast cancer are very complex, but recent research showed that weight gain was associated with the risk of breast cancer. Medical researchers (Huang et al., 1997) studied over 95,000 U.S. female nurses aged 30 to 55. The nurses were followed for 16 years. They discovered that weight gain after the age of 18 years was unrelated to breast cancer incidence before menopause, but was associated with it after menopause. Post-menopausal weight gain increased both the risk of breast cancer and mortality. About 16 percent of the breast cancers were attributed to excessive weight gain ( $\geq 20.1$  kg). Avoiding excessive weight gain during adulthood appears to be an important factor in reducing a woman's risk of breast cancer.

## Strength and Injury

Industrial musculoskeletal injury is a major problem associated with physically demanding jobs. Of all manual materials handling tasks, lifting accounts for about 50 percent of these injuries (Ayoub, 1997; Snook et al., 1978). Nearly 25 percent of all worker compensation claims in the United States are related to low-back injuries, with an economic effect estimated to be as high as \$20 billion (Ayoub, 1997). It is believed by some that the lack of strength increases the risk of injury associated with industrial manual materials handling tasks.

Ergonomic research suggests that it is not only the strength of the worker that is involved, but also the demands of the task. In a major retrospective study, Bigos and associates (Bigos et al., 1986) reported that industrial musculoskeletal injury was not related to muscular strength. The injury rates of stronger workers did not differ from their weaker counterparts. Missing in their study was data on the demands of the work task. In contrast, ergonomic research (Dehlin et al., 1976; Herrin, 1986; Magora, 1970; Snook et al., 1978) has established a relationship between industrial back injuries and psychophysical lift capacity. Psychophysics quantify the demand of physical tasks by relative intensity or a percentage of maximum capacity (Borg, 1998; Resnik, 1995). Both cross-sectional (Herrin, 1986; Liles, 1984) and longitudinal (Chaffin, 1974; Chaffin, Herrin, & Keyserling, 1978) research showed that the risk of industrial back injury increased exponentially as the force required to complete the lift approached the employee's maximum isometric strength capacity. In a classic, epidemiological retrospective study, Snook and Associates (Snook et al., 1978) showed that workers were three times more susceptible to low-back injury if they lifted weight loads psychophysically judged to be too heavy for 75 percent of the industrial population.

A reason for including a back flexibility item in an adult fitness test (Golding et al., 1989) is that the lack of low back flexibility is believed to be a risk factor of low back pain. Plowman (Plowman, 1992) published a comprehensive review of the research relating flexibility and low back pain. This review did not show that the lack of flexibility was a risk factor of low back pain. A review of the literature failed to provide evidence that flexibility was a valid predictor of physically demanding work performance.

---

## Summary

This chapter reviewed the types of physical performance tests that can be used for evaluating a worker's capacity to do physically demanding work. In addition, a comprehensive discussion was provided of the methods that can be used to measure aerobic fitness, body composition, strength, muscular endurance, and flexibility. The tests most commonly used in the public sector include strength and  $\text{VO}_2\text{max}$  tests. Strength tests have been shown to be significantly correlated with many materials handling tasks, and  $\text{VO}_2\text{max}$  is correlated with endurance tasks such as firefighting. Military research has examined the role of body composition on materials handling tasks and showed that fat-free weight is related to these tasks. Work sample tests duplicate or simulate critical work tasks. Although work sample tests can have the advantage of being content valid, the lim-



itations are that they can increase the risk of injury and do not provide information about the applicant's maximum work capacity. Another potential limitation of work sample tests is that there can be a factor of skill required by the test. Although an applicant may not have the skill to effectively perform the task, with practice the skill can be mastered. A promising approach is to use basic ability tests to define an applicant's physiological capacity to meet the demands of the task. This strategy is discussed in more detail in the validity chapter. More research is needed to define the physiological determinates of physically demanding work tasks. The final section of the chapter provides a brief overview of the role of aerobic fitness, body composition, and strength on health and injury. Epidemiological research documents that low aerobic fitness and unfavorable body composition are associated with higher levels of degenerative diseases such as hypertension, stroke, heart disease, and diabetes. Research has shown that the risk of injury is a function of the combined effect of worker strength and demands of the task. The risk of injury increases as the demands of the task approach the physiological capacity of the worker.

## Endnotes

1. Many people use the term lean body mass or lean body weight instead of fat-free weight. Lean body weight has a density of less than 1.100 g/cc because it contains from 2 to 3 percent essential lipid. Lohman (1992) maintains that fat-free weight is the proper term.
2. These investigators included Dr. Andrew S. Jackson, Department of Health and Human Performance, University of Houston; Dr. Hobart Osburn, Department of Psychology, University of Houston; and Dr. Kenneth R. Laughery, Department of Psychology, Rice University.
3. The raw score work sample test distribution was skewed. A log transform was used to more closely approximate a normal distribution.

---

## References

1. American College of Sports Medicine (ACSM). (1990). The recommended quantity and quality of exercise for developing and maintaining cardiorespiratory and muscular fitness in healthy adults. *Medicine and Science in Sports and Exercise*, 22, 265–274.
2. American College of Sports Medicine (ACSM). (1991). Guidelines for exercise testing and prescription. (3<sup>rd</sup> ed.). (Vol. 4). Philadelphia: Lea and Febiger.
3. Arnold, J. D., Rauschenberger, J. M., Soubel, W. G., & Guion, R. M. (1982). Validation and utility of a strength test for selecting steelworkers. *Journal of Applied Psychology*, 67, 588–604.
4. Arvey, R. D., Nutting, S. M., & Landon, T. E. (1992). Validation strategies for physical ability testing in police and fire settings. *Public Personnel Management*, 21, 301–312.
5. Astrand, P., & Rodahl, K. (1970). *Textbook of work physiology*. New York: McGraw-Hill.
6. Astrand, P.-O., & Rodahl, K. (1986). *Textbook of work physiology*. (3<sup>rd</sup> ed.). New York: McGraw-Hill.
7. Astrand, P.-O., & Ryhming, I. (1954). A nomogram for calculation of aerobic capacity (physical fitness) from pulse rate during submaximal work. *Journal of Applied Physiology*, 7, 218–221.

8. Ayoub, M. A. (1982). Control of manual lifting hazards: III. Preemployment screening. *Journal Of Occupational Medicine*, 24, 751–761.
9. Ayoub, M. M., Dempsey, P. G., & Karwowski, W. (1997). Chapter 34: Manual materials handling. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (pp. 1085–1123). New York: John Wiley & Sons, Inc.
10. Bar-Or, O. (1987). The Wingate anaerobic test: An update on methodology, reliability and validity. *Sports Medicine*, 4, 381–394.
11. Barnard, R., & Duncan, H. W. (1975). Heart rate and ECG responses of firefighters. *Journal Of Occupational Medicine*, 17, 247–250.
12. Baumgartner, T. A., & Jackson, A. S. (1999). *Measurement for evaluation in physical education and exercise science*. (6<sup>th</sup> ed.). Dubuque: Wm. C. Brown.
13. Bernauer, E. M., & Bonanno, J. (1975). Development of physical profiles for specific jobs. *Journal of Occupational Medicine*, 17, 22–33.
14. Bigos, S., Spengler, D., Martin, N., Zeh, J., Fisher, L., & Nachemson, A. (1986). Back injuries in industry: A retrospective study III. *Spine*, 3, 252–256.
15. Blair, S. N., III, H. W. K., Paffenbarger Jr., R. S., Clark, D. G., Cooper, K. H., & Gibbons, L. W. (1989). Physical fitness and all-cause mortality: A prospective study of health men and women. *Journal of the American Medical Association*, 262, 2395–2401.
16. Blair, S. N., Kohl III, H. W., Barlow, M. S., Paffenbarger Jr., R. S., Gibbons, L. W., & Macera, C. A. (1995). Changes in physical fitness and all-cause mortality: A prospective study of healthy and unhealthy men. *Journal of the American Medical Association*, 273, (14), 1093–1098.
17. Borg, G. (1998). *Borg's perceived exertion and pain scaling method*. Champaign: Human Kinetics.
18. Bouchard, C. L., & Blair, S. N. (1999). Introductory comments for the consensus on physical activity and obesity. *Medicine and Science in Sports and Exercise*, 31, S498–S501.
19. Bray, G. A. (1993). Fat distribution and body weight. *Obesity Research*, 1, 203–205.
20. Brozek, J., Grande, F., & Anderson, J. T. (1963). Densitometric analysis of body composition: Revision of some quantitative assumptions. *Annals of New York Academy of Science*, 110, 113–140.
21. Bruce, R. A., Kusumi, F., & Hosmer, D. (1973). Maximal oxygen intake and nomographic assessment of functional aerobic impairment in cardiovascular disease. *American Heart Journal* 85, 546–562.
22. Buskirk, E. R., & Hodgson, J. L. (1987). Age and aerobic power: the rate of change in men and women. *Federation Proceedings*, 46, 1824–1829.
23. Chaffin, D. B. (1974). Human strength capability and low-back pain. *Journal of Occupational Medicine*, 16, 248–254.
24. Chaffin, D. B. (1975). Ergonomics guide for the assessment of human static strength. *American Industrial Hygiene Association Journal* 36, 505–511.
25. Chaffin, D. B., Herrin, G. D., & Keyserling, W. M. (1978). Preemployment strength testing. *Journal of Occupational Medicine*, 67, 403–408.
26. Cooper, K. H. (1968). A means of assessing maximal oxygen intake. *Journal of the American Medical Association*, 203, 201–204.
27. Cureton, K. J. (1984). A reaction to the manuscript by Jackson. *Medicine and Science in Sports and Exercise*, 16, 621–622.

28. Cureton, K. J., Sloniger, M. A., O'Bannon, J. P., Black, D. M., & McCoormack, W. P. (1995). A generalized equation for prediction of  $\text{VO}_2$  peak from 1-mile run/walk performance. *Medicine and Science in Sports and Exercise*, 27, (3), 445–451.
29. Davis, P. O., Dotson, C. O., O'Connor, J. S., & Confessore, R. J. (1992). *The development of a job-related physical performance test for Saint Paul firefighters*. Saint Paul, MN: Saint Paul Fire Department.
30. Dehlin, O., Hendenrud, B., & Horal, J. (1976). Back symptoms in nursing aids in a geriatric hospital. *Scandinavian Journal Of Rehabilitative Medicine*, 8, 47–53.
31. Dempsey, P. G., Ayoub, M. M., Westfall, P. H. (1998). Evaluation of the ability of power to predict low frequency lifting capacity. *Ergonomics*, 41, 1222–1241.
32. deVries, H. A., Housh, T. J. (1994). *Physiology of exercise for physical education, athletics and exercise science*. (5<sup>th</sup> ed.). Dubuque, IA: Wm. C. Brown.
33. Durnin, J. V. G. A., & Passmore, R. (1967). *Energy, work and leisure*. London: Heinemann Educational Books, LTD.
34. Durnin, J. V. G. A., & Rahaman, M. M. (1967). The assessment of the amount of fat in the human body from measurements of skinfold thickness. *British Journal of Nutrition*, 21, 681–689.
35. Foster, C., Jackson, A. S., & Pollock, M. L. (1984). Generalized equations for predicting functional capacity from treadmill performance. *American Heart Journal*, 107, 1229–1234.
36. Gallagher, D., Visser, M., Sepulveda, D., Pierson, R. N., Harris, T., Heymsfield, S. B. (1996). How useful is body mass index for comparison of body fatness across age, sex, and ethnic groups. *American Journal Of Epidemiology*, 143, (3), 228–239.
37. Gettman, L. R. (1993). *Chapter 19 Fitness Testing* In ACSM, Resource Manual for Guidelines for Exercise Testing and Prescription (2<sup>nd</sup> ed., pp. 229–246). Philadelphia: Lea & Febiger.
38. Golding, L. A., Meyers, C. R., & Sinning, W. E. (1989). *The Y's Way to Physical Fitness*. (3<sup>rd</sup> ed.). Chicago, IL: National Board of YMCA.
39. Harris, M. (1969). A factor analytic study of flexibility. *Research Quarterly*, 40, 62–70.
40. Herrin, G. D., Jaraiedi, M., & Anderson, C. K. (1986). Prediction of overexertion injuries using biomechanical and psychophysical models. *American Industrial Hygiene Association Journal*, 47, 322–330.
41. Heymsfield, S. B., Wang, Z., & Withers, R. T. (1996). Chapter 7 Multicomponent molecular level models of body composition analysis. In A. F. Roche, Heymsfield, S. B., Lohman, T. G. (Eds.), *Human Body Composition* (pp. 129–148). Champaign: Human Kinetics.
42. Hodgdon, J. A. (1992). Body composition in the military services: Standards and methods. In B. M. Marriott & J. Grumstrup-Scott (Eds.), *Body Composition and Physical Performance: Applications for the Military Services* (pp. 57–70). Washington, DC: National Academy Press.
43. Hodgdon, J. A., & Beckett, M. B. (1984a). *Prediction of percent body fat for US Navy Men from body circumferences and height*. San Diego, CA: Report No. 84–11, Naval Health Research Center.
44. Hodgdon, J. A., & Beckett, M. B. (1984b). *Technique for measuring body circumferences and skinfold thickness*. San Diego, CA: Report No. 84–39, Naval Health Research Center.
45. Hogan, J. C. (1991). Chapter 11 Physical Abilities. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of Industrial and Organizational Psychology* (2<sup>nd</sup> ed., Vol. 2, pp. 743–831). Palo Alto: Consulting Psychologist Press, Inc.
46. Howley, E. T., Bassett, D. R. (1997). Maximal oxygen uptake: Classical versus contemporary viewpoints. *Medicine and Science in Sports and Exercise*, 29, 591–603.

47. Huang, Z. et al. (1997). Dual effects of weight and weight gain on breast cancer risk. *Journal of the American Medical Association*, 278, (17), 1407–1411.
48. Hubert, H. B., et al. (1983). Obesity as an independent risk factor for cardiovascular diseases: A 26-year follow-up of participants in the Framingham heart study. *Circulation*, 67, 968–977.
49. Jackson, A. S. (1984). Research design and analysis of data procedures for predicting body density. *Medicine and Science in Sports and Exercise*, 16, 616–620.
50. Jackson, A. S. (1994). Chapter 3 Preemployment Physical Evaluation. *Exercise and Sport Science Reviews*, 22, 53–90.
51. Jackson, A. S., Sekula, B. K. (1999). The influence of strength and gender on defining psychophysical lift capacity. *Proceeding of the Human Factors and Ergonomics Society*, 43, 723–727.
52. Jackson, A. S., Beard, E. F., Wier, L. T., Ross, R. M., Stuteville, J. E., & Blair, S. N. (1995). Changes in aerobic power of men ages 25–70 years. *Medicine and Science in Sports and Exercise*, 27, 113–120.
53. Jackson, A. S., Blair, S. N., Mahar, M. T., Wier, L. T., Ross, R. M., & Stuteville, J. E. (1990). Prediction of functional aerobic capacity without exercise testing. *Medicine and Science in Sports and Exercise*, 22, 863–870.
54. Jackson, A. S., Borg, G., Zhang, J. J., Laughery, K. R., & Chen, J. (1996a). Function of physical working capacity on psychophysical lift ratings. July, 1996, pp. 309–314. In A. Mital, H. Krueger, S. Kumar, M. Menozzi, & J. Fernandez (Eds.), *Advances in Occupational Ergonomics and Safety I Proceedings of the XI<sup>th</sup> Annual International Occupational Ergonomics and Safety Conference*. Zurich, Switzerland; International Society for Occupational Ergonomics and Safety.
55. Jackson, A. S., Borg, G., Zhang, J. J., Laughery, K. R., & Chen, J. (1997). Role of physical work capacity and load weight on psychophysical lift ratings. *International Journal of Industrial Ergonomics*, 20, 181–190.
56. Jackson, A. S., Osburn, H. G., & Laughery, K. R. (1984). Validity of isometric strength tests for predicting performance in physically demanding jobs. *Proceedings of the Human Factors Society 28<sup>th</sup> Annual Meeting*, 28, 452–454.
57. Jackson, A. S., Osburn, H. G., Laughery, K. R., & Vaubel, K. P. (1992). Validity of isometric strength tests for predicting the capacity to crack, open and close industrial valves. *Proceedings of the Human Factors Society 36<sup>th</sup> Annual Meeting*, 1, 688–691.
58. Jackson, A. S., Osburn, H. G., Laughery, K. R., & Young, S. L. (1993). *Validation of Physical Strength Tests for the Federal Express Corporation*. Houston: Center of Applied Psychological Services, Rice University.
59. Jackson, A. S., Osburn, H. G., & Laughery, Sr., K. R. (1991a). Validity of isometric strength tests for predicting endurance work tasks of coal miners. *Proceedings of the Human Factors Society 35<sup>th</sup> Annual Meeting*, 1, 763–767.
60. Jackson, A. S., Osburn, H. G., Laughery Sr., K. R., & Vaubel, K. P. (1991b). Strength demands of chemical plant work tasks. *Proceedings of the Human Factors Society 35<sup>th</sup> Annual Meeting*, 1, 758–762.
61. Jackson, A. S., & Pollock, M. L. (1976). Factor analysis and multivariate scaling of anthropometric variables for the assessment of body composition. *Medicine and Science in Sports*, 8, 196–203.
62. Jackson, A. S., & Pollock, M. L. (1978). Generalized equations for predicting body density of men. *British Journal of Nutrition*, 40, 497–504.
63. Jackson, A. S., Pollock, M. L., Graves, J. L., & Mahar, M. T. (1988). Reliability and validity of bio-electrical impedance in determining body composition. *Journal of Applied Physiology*, 64, 529–534.

64. Jackson, A. S., Pollock, M. L., & Ward, A. (1980). Generalized equations for predicting body density of women, *Medicine and Science in Sports and Exercise*, 12, 175–182.
65. Jackson, A. S., Wier, L. T., Ayers, G. W., Beard, E. F., Stuteville, J. E., & Blair, S. N. (1996b). Changes in aerobic power of women, ages 20 to 64 years. *Medicine and Science in Sports and Exercise*, 28, 884–891.
66. Jeanneret & Associates, I. (1999). *Evaluation of physical ability and written tests for entry-level fire-fighters*. St. Paul, MN: City of St. Paul Fire Department.
67. Kline, G. M., Porcari, J. P., Hintermeister, R., Freedson, P. S., Ward, A., McCarron, R. F., Ross, J., & Rippe, J. M. (1987). Estimation of  $\text{VO}_2\text{max}$  from a one-mile track walk, gender, age, and body weight. *Medicine and Science in Sports and Exercise*, 19, 253–259.
68. Laughery K. R. & Jackson, A. S. (1985). *Physical abilities and performance tests for coal miner jobs*. Houston, TX Center for Psychological Services, Rice University.
69. Laughlin, M. S. (1998). *The relationship between isometric and isotonic closed kinetic chain leg strength*. Unpublished M.S., University of Houston, Houston.
70. Lee, C. D., Blair, S. N., Jackson, A. S. (1999). Cardiorespiratory fitness, body composition, and all-cause and cardiovascular disease mortality in men. *American Journal of Clinical Nutrition*, 69, 373–380.
71. Lemon, P., & Hermiston, R. (1977a). Physiological profile of professional fire fighters. *Journal of Occupational Medicine*, 19, 337–340.
72. Lemon, P., & Hermiston, R. T. (1977b). The human energy cost of firefighting. *Journal of Occupational Medicine*, 19, 558–562.
73. Lew, E. A., & Garfinkel, L. (1979). Variations in mortality by weight among 750,000 men and women. *Journal of Chronic Diseases*, 32, 181–225.
74. Liles, D. H., Deivanayagam, S., Ayoub, M. M., & Mahajan, P. (1984). A job severity index for the evaluation and control of lifting injury. *Human Factors*, 26, 683–693.
75. Lohman, T. G. (1982). Body composition methodology in sport medicine. *Physician and Sports Medicine*, 10, 46–58.
76. Lohman, T. G. (1992). *Advances in body composition assessment*. Champaign: Human Kinetics Publishers.
77. Lohman, T. G. (1996). Chapter 4 Dual Energy X-Ray Absorptiometry. In A. F. Roche, Heymsfield, S. B., Lohman, T. G. (Eds.), *Human body composition* (pp. 3–23). Champaign: Human Kinetics.
78. Magora, A. (1970). Investigation of the relation between low back pain and occupation. *Industrial Medicine and Surgery*, 39, 504–510.
79. Mahar, M., Jackson, A., & Ross, R. (1985). Predictive accuracy of single and double stage submax treadmill work for estimating aerobic capacity. *Medicine and Science in Sports and Exercise*, 17, 206–207.
80. Manning, J., & Griggs, T. (1983). Heart rates in fire fighters using light and heavy breathing equipment: similar near-maximal exertion in response to multiple work load conditions. *Journal of Occupational Medicine*, 25, 215–218.
81. McArdle, W. D., Katch, F. I., & Katch, V. L. (1991). *Exercise physiology: Energy, nutrition, and human performance*. (3rd ed.). Philadelphia: Lea & Febiger.
82. Mitchell, J. H., & Blomqvist, G. (1971). Maximal oxygen uptake. *The New England Journal of Medicine*, 284, 1018–1022.

83. Mitchell, J. H., Sproule, B. J., & Chapman, C. B. (1958). The physiological meaning of the maximal oxygen intake test. *Journal of Clinical Investigation*, 37, 538–547.
84. Morrow, J. R., Jackson, A. S., Bradley, P. W., & Hartung, G. H. (1986). Accuracy of measured and predicted residual lung volume on body density measurement. *Medicine and Science in Sport and Exercise*, 18, 647–652.
85. NIOSH. (1977). *Preemployment strength testing*. Washington, DC: U.S. Department of Health and Human Services.
86. Noakes, T. D. (1988). Implications of exercise testing for prediction of athletic performance: A contemporary perspective. *Medicine and Science in Sports and Exercise*, 20, 319–330.
87. Noakes, T. D. (1997). Challenging beliefs: Ex Africa semper aliquid novi. *Medicine and Science in Sports and Exercise*, 29, 571–590.
88. Noakes, T. D. (1998). Maximal oxygen uptake: “classical” versus “contemporary” viewpoints: A rebuttal. *Medicine and Science in Sports and Exercise*, 30, 1381–1398.
89. O’Connell, E., Thomas, P., Caddy, L., & Karwasky, R. (1986). Energy costs of simulated stair climbing as a job-related task in fire fighting. *Journal of Occupational Medicine*, 28, 282–284.
90. Plowman, S. A. (1992). Chapter 8 Physical activity, physical fitness, and low back pain. In J. O. Holloszy (Ed.), *Exercise and Sport Sciences Reviews* (Vol. 20, pp. 221–242). Baltimore, MD: Williams & Wilkins.
91. Pollock, M. L., Hickman, T., & Kendrick, Z. (1976). Prediction of body density in young and middle-aged men. *Journal of Applied Physiology*, 40, 300–304.
92. Reilly, R. R., Zedeck, S., & Tenopyr, M. L. (1979). Validity and fairness of physical ability tests for predicting craft jobs. *Journal of Applied Psychology*, 64, 267–274.
93. Resnik, M. L. (1995). The generalizability of psychophysical ratings in predicting the perception of lift difficulty. *Proceedings of the Human Factors Society*, 679–682.
94. Ross, R. M., & Jackson, A. S. (1986). Development and validation of total-work equations for estimating the energy cost of walking. *Journal of Cardiac Rehabilitation*, 6, 185–192.
95. Ross, R. M., & Jackson, A. S. (1990). *Exercise concepts, calculations, and computer applications*. Carmel, IN: Benchmark Press.
96. Rowell, L. B., Taylor, H. L., & Wang, Y. (1964). Limitations to prediction of maximal oxygen intake. *Journal of Applied Physiology*, 19, 919–927.
97. SAS. (1998). *Statview* (Version 5.0). Cary, NC: SAS Institute, Inc.
98. Schutte, J. E., Townsend, E. J., Hugg, J., Shoup, R. F., Malina, R. M., & Blomqvist, C. G. (1984). Density of lean body mass is greater in blacks than in whites. *Journal of Applied Physiology: Respiratory, Environmental and Exercise Physiology*, 56, 1647–1649.
99. Sinning, W. E. (1996). Chapter 13. Body Composition in Athletes. In A. F. Roche, Heymsfield, S. B., Lohman, T. G. (Eds.), *Human Body Composition*. Champaign, IL: Human Kinetics.
100. Siri, W. E. (1961). Body composition from fluid space and density. In J. Brozek and A. Hanschel, (Eds.), *Techniques for measuring body composition*. Washington, DC: National Academy of Science.
101. Snook, S. H., Campanelli, R. A., & Hart, J. W. (1978). A study of three preventive approaches to low back injury. *Journal of Occupational Medicine*, 20, 478–481.
102. Snook, S. H., & Ciriello, V. M. (1991). The design of manual handling tasks: Revised tables of maximum acceptable weights and forces. *Ergonomics*, 34, 1197–1213.

103. Sothmann, M. S., Saupe, K., Jasenor, D., & Blaney, J. (1992). Heart rate response of firefighters to actual emergencies. *Journal of Occupational Medicine*, 34, 797–800.
104. Sothmann, M. S., Saupe, K. W., Jasenor, D., Blaney, J., Donahue-Fuhrman, S., & Woulfe, T. (1990). Advancing age and the cardiorespiratory stress of fire suppression: Determining a minimum standard for aerobic fitness. *Human Performance*, 3, 217–236.
105. Tran, Z., & Weltman, A. (1989). Generalized equation for predicting body density of women from girth measurements. *Medicine and Science in Sports and Exercise*, 21, 101–104.
106. Tran, Z. V., Weltman, A., & Seip, R. L. (1988). Predicting body composition of men from girth measurements. *Human Biology*, 60, 167–176.
107. USDHHS. (1990). *Healthy people 2000: National health promotion and disease prevention objectives*. (DHHS Publication No. (PHS) 91–50212 ed.). Washington DC: Department of Health and Human Services.
108. VanLoan, M. D. (1996). Chapter 11. Total body composition—birth to old age. In A. F. Roche, Heymsfield, S.B., Lohman, T.G. (Ed.), *Human Body Composition*. Champaign, IL: Human Kinetics.
109. Vickery, S. R., Cureton, K.J., Collins, M. A. (1988). Prediction of body density from skinfolds in black and white young men. *Human Biology*, 60, 135–149.
110. Vogel, J. A., & Friedl, K. E. (1992). Army Data: Body composition and physical capacity. In B. M. Marriott & J. Grumstrup-Scott (Eds.), *Body composition and physical performance: Applications for the military services* (pp. 89–104). Washington, D.C.: National Academy Press.
111. Wilmore, J. H., Despres, Jean-pierre, Stanforth, P. R., Mandel, S., Rice, T., Gagnon, J., Leon, A. S., Rao, D. C., Skinner, J. S., Bouchard, C. (1999). Alterations in body weight and composition consequent to 20 week of endurance training: The HERITAGE family study. *American Journal of Clinical Nutrition*, 70, 346–352.
112. Wilmore, J. H., & Costill, D. L. (1994). *Physiology of sport and exercise*. Champaign, IL: Human Kinetics.





# Chapter 5

## Physical Test Validation for Job Selection

**James A. Hodgdon, Ph.D.**

Human Performance Center  
Naval Health Research Center  
San Diego, CA

**Andrew S. Jackson, P.E.D.**

Department of Health and Human Performance  
University of Houston  
Houston, TX

---

### Abstract

This chapter examines the issues related to physical test validation for job selection. The chapter is divided into three major sections. The first examines issues and accepted methods of test validation. The focus is on the interpretation of the Equal Employment Opportunity Commission (EEOC) guidelines (EEOC, 1978) as they relate to test validation. The sanctioned validation methods are content validity, criterion-related validity, and construct validity. The measurement theory used to evaluate the quality of employment tests is based on the American Psychological Association standards for validating educational and psychological tests (A.P.A., 1985; A.P.A., 1987). A major difference in physical test validation is the use of physiological rather than psychological tests. The second section of the chapter examines the differences between physiological and psychological test validation. The goal of physiological validation is to define the physiological capacity needed by a worker to perform the work demanded by the task. Principal features of the physiological validation approach are the use of a physiological metric to quantify test performance and the interpretation of validity results with relevant physiological research and theory. The final section of the chapter reviews published employment validation research on physical tests.

---

### Employment Selection Tests

The principal guidance for the design and implementation of selection tests for employment is the Uniform Guidelines on Employee Selection Procedures issued by the Equal Employment Opportunity Commission in 1978 (EEOC, 1978). These guidelines state that a selection procedure

has “adverse impact” if the selection rate for any group is less than 80 percent for the group with the highest selection rate. Selection procedures that have adverse impact are considered discriminatory unless they can be justified. A selection procedure that has adverse impact can be justified if—

1. The tests or measures are derived from a job analysis
2. The tests or measures are indicators of critical or important job duties, work behaviors, or work outcomes
3. The tests or measures have been shown to be valid indicators of such duties, behaviors, or outcomes.

This existence of a procedure for justifying selection tests is critical in the area of selection based on physical abilities. There are well-recognized differences in physical abilities between genders (McArdle, Katch, & Katch, 1996), and the development of a physical abilities selection test for physically demanding jobs runs a great risk of having adverse impact across gender.

The nature of job analyses, identification of critical or important job duties, and nature of physical selection tests are discussed in other sections. This section considers issues surrounding the demonstration of the validity of selection tests or measures. As in other sections of this report, the emphasis is on selection based on physical ability.

## Validity of Selection Tests

The extent to which a test or set of tests measures what it is meant to measure is called the validity of the test. For the purposes of this chapter, validity is the accuracy with which selection test(s) measure important work behaviors (Jackson, 1994). The Uniform Guidelines recognize three types of validity with respect to selection test development: content validity, criterion-related validity, and construct validity.

**Content Validity**—That a test has content validity means that the test items reflect important elements of the job. The job and test content are linked. Most content-valid test items are, in fact, job samples or simulations of job tasks. Theoretically, for the test as a whole to be content valid, the test items must sample all critical or important duties, work behaviors, or work outcomes. For example, if a job has two critical, physically demanding tasks, one involving repeated lifting to a fixed height and one involving carrying materials a long distance, both tasks need to be simulated in the content-valid selection test. Such job sample tasks are usually scored as to whether the applicant can or cannot perform the task. Additionally, for jobs that have time constraints, such as emergency service tasks, there may be time limits imposed for task completion. Successful completion of the tasks qualifies one for the job. Content-valid tests are the most defensible tests because they are the most direct indicators of job performance capability. The closer the simulation is to the actual job task, the more defensible it is as a selection test.

**Criterion-Related Validity**—A test is said to have criterion-related validity when the test items are shown to be estimators or predictors of critical or important duties, work behaviors, or work outcomes. Criterion-related validity is usually expressed as a correlation coefficient between test per-

formance (the predictor) and performance of an important or critical job element or behavior (the criterion). The criterion job element can be any of a number of job behaviors including work-task performance, injury rates on the job, absenteeism, or peer or supervisor ratings. Criterion-related selection tests are not, by definition, direct indicators of the ability to perform a job or job task. They rely on a secondary relationship between the criterion task and the predictor test.

Two types of criterion-related validity can be distinguished. A test is said to have concurrent validity whenever the test is used to predict a current capability. An example is use of a bench press 1-repetition maximum (1RM) is used to predict an applicant's current ability to lift a 50-kg box to elbow height. If the test is used to predict some future event, it is said to have predictive validity. An example would be the use of the time to complete a 1-mile run as an indicator of future success in a Military training program.

Correlational studies are carried out to demonstrate criterion-related validity. Critical or important job behaviors are determined during the job analysis. The nature of the critical job behaviors usually suggests the nature of the selection test to be employed. If a critical task requires lifting, for example, then selection tests that measure strength would be appropriate. If the critical job task requires prolonged activity, then a test related to endurance, such as a run for time, might be appropriate. Once candidate tests have been chosen, the tests are administered to a sample of workers or another suitable sample. Their performance on the identified critical job tasks (or other criterion measures) is also measured. The strength of the associations between performance on the selection tests and performance on the critical job behaviors is expressed as the correlation coefficient, which is a measure of the amount of common variance accounted for by two measures. If the correlation coefficient between a selection test performance and performance on a critical job behavior is suitably high, the selection test may be used. It should be noted that there is no standard for the minimum acceptable correlation coefficient between a selection test and job behavior. Statistical significance is not always a good indicator because with large sample sizes, a correlation that explains only a small part of the variance can be significant. That which is possible or practical may drive the selection of an acceptable level of correlation. As a benchmark, one might note that a correlation coefficient of 0.707 indicates that 50 percent of the common variance in the relationship has been explained, but this is difficult to use as a criterion because many things can affect the size of a correlation coefficient. For example, the size of correlation is influenced substantially by the variability of the sample tested. It is also possible to have a high correlation but considerable errors in prediction (Altman & Bland, 1983; Altman & Bland, 1986). This subject is covered in more detail in another section of this chapter.

The scoring of criterion-related tests is based on the achievement of critical performance levels on the selection test(s). These critical performance levels can be quite difficult to define. Usually, they are derived from a mathematical function relating the predictor and criterion performances. The value of the performance on the selection test that is associated mathematically with a critical level of performance on the important job task is used as the cut off score or cut-score on the selection test. This critical level of job performance needs to be identified in the job analysis. This subject is covered in more detail in another section of this chapter.

Even in the simplest case, when a single critical task and critical level of performance, and a single predictor measure are identified, it can be difficult to set a critical level of performance. This is because the relationship between performance on the selection test and performance on the crite-

tion task is not perfect. As an example, Figure 5.1 shows the relationship between the maximum weight box that can be lifted to elbow height (the work tasks) and 1RM for arm-curl (the criterion test). As one can see, arm-curl 1RM and maximum box weight appear to be strongly related. The correlation coefficient for this relationship is 0.875. Furthermore, the relationship between the variables appears to be a straight line, as suggested by the diagonal line crossing the figure. This line represents the linear regression of maximal box lift weight with arm-curl 1RM. However, the points are scattered about the line. If the critical task for a particular job involved lifting a 50-kg box to elbow height (the value indicated by the horizontal line), the mean arm-curl value associated with this box weight is 23.4 kg (the solid vertical line). This, ideally, would be the critical arm-

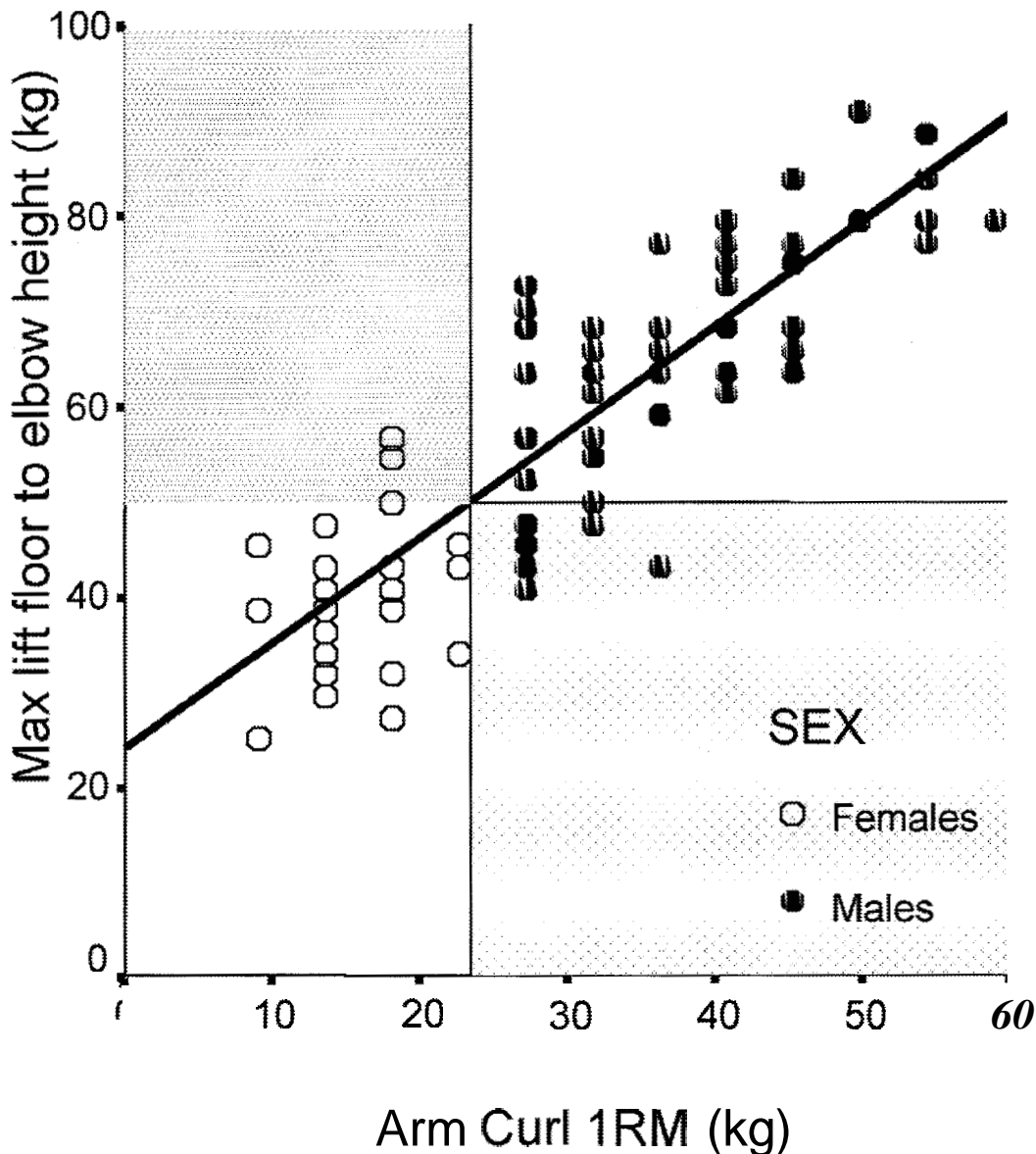


Figure 5.1 Maximum box weight lifted to elbow height as a function of arm curl 1RM

curl 1RM value that we would pick if arm-curl 1RM were the selection task for this job. However, it is clear by inspection of Figure 5.1 that some individuals who lifted less than 23.4 kg on the arm-curl could lift a 50-kg box. These individuals are called “false negatives” because they failed the test (and are not selected) but can perform the work task. In Figure 5.1, the false negatives appear in the upper left quadrant formed by the horizontal and vertical lines within the figure. Similarly, some individuals who lifted more than 23.4 kg could not lift a 50-kg box. These individuals are known as “false positives” because they passed the test (and were selected), but cannot perform the work task. In Figure 5.1, these individuals appear in the lower right quadrant. The Uniform Guidelines allow the exercise of a certain amount of judgment in setting cut-scores. However, one needs to have a defensible rationale. These issues are examined in more detail in the physiological validation section of this chapter.

**Construct Validity**—Construct validity is the most indirect and theory-driven method of establishing validity. Construct validity exists when selection tests are related to a general trait or set of characteristics (the construct) that is associated with successful accomplishment of important or critical job behaviors. The establishment of construct validity requires that employers show that a construct (a general trait or set of characteristics) is required for satisfactory job performance, and that the selection test or tests measure this same construct.

Constructs are often developed using the statistical technique of factor analysis (Rummel, 1970). In factor analysis, a number of correlated variables are reduced to a smaller number of dimensions or factors. Within the factor, each of the included variables has a coefficient or “loading,” a numerical value indicating the strength of association of that variable with the factor. The greater the loading, the greater the association between the variable and the factor. The factor is defined mathematically as the sum of the factor variable values, each multiplied by its loading. The variables with the greatest loadings drive the theoretical interpretation of the factor.

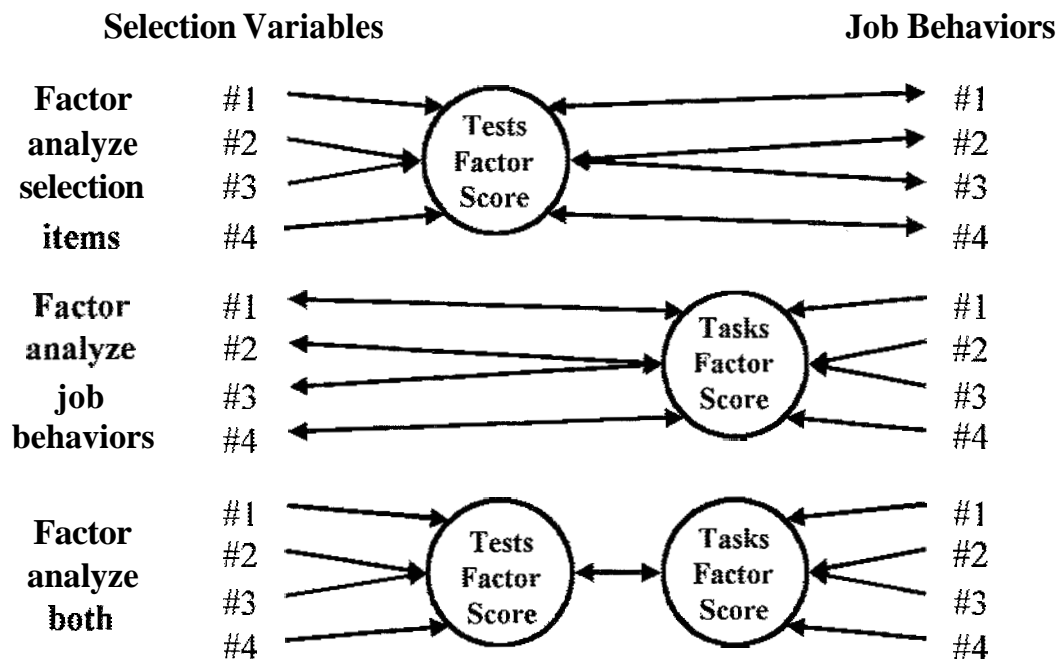
Construct validity can be established in three ways—

1. Performances on job behaviors can be analyzed to determine dimensions within the job. Scores on selection tests can then be shown to be correlated with the job dimensions.
2. Scores on selection tests can be factor analyzed, and dimensions within the selection tests identified. A number of examples of such analyses can be found in the literature (Fleishman, 1964; Hogan, 1991a; Meyers, Gebhardt, Crump, & Fleishman, 1984)
3. Factor scores from the dimensions of the selection tests can be shown to be correlated to performance on important job behaviors. Both potential selection test items and performance on important job behaviors can be factor analyzed. A validity study can then be carried out to analyze the associations between the selection factors and the job factors.

These options are indicated schematically in Figure 5.2.

Figure 5.2 is an oversimplified version of the actual situation. Often, more than one construct is present in the job behaviors. For example, strength and endurance may be required for job success. In such a case, many more relationships must be worked out in the validity study.

The conduct of a study to demonstrate construct validity is similar to that for criterion-related validity except that instead of a one-to-one mapping of performance on a selection test to perform-



**Figure 5.2** Three experimental designs for construct validity studies. Single-ended arrows indicate variables included in the factor analysis. Double-ended arrows indicate correlations to be measured.

ance on a job behavior, several selection-test items are measured that are used to calculate factor scores to represent the selection constructs being measured, and/or several job behaviors are measured to calculate factor scores to represent the job constructs being measured. It is these factor scores that are used in the correlational analysis. Construct-validity relationships are often difficult to demonstrate because of the need to identify the factor structures in the job and selection tests and then establish associations between or among them. Given these difficulties, many employers choose to use the measures of underlying constructs directly as elements of criterion-related validity studies.

## Requirements for Validity Studies

The Uniform Guidelines provide general and technical standards for validity studies. Among the general standards are the following—

- \* In addition to specifying the three types of studies (content, criterion-related, and construct-validity), the guidelines require the studies to be consistent with applicable professional standards for such research, accurate and free from bias.
- The validity studies should be documented.
- The employer must be prepared to justify the method used to implement the selection tests. If use of a test has greater adverse impact when used as a ranking device than if it were

implemented as a simple pass/fail, then the employer must provide sufficient evidence of the validity and utility to support use of the test to rank-order participants.

- Selection procedures may be developed for higher level jobs in cases where most of the entry-level applicants will progress to those higher level jobs.
- **An** employer may continue to use selection procedures for which there is not yet full validity evidence as long as the employer has evidence of the substantial validity of the procedures and will conduct, when technically feasible, a study to produce the additional evidence required.
- Employers may also use validity studies conducted by others when it can be shown that the validity studies were conducted properly and that the jobs perform substantially the same major work behaviors for the employer as for those who conducted the study.
- Employers, labor organizations, and employment agencies are encouraged to work together and cooperate in validity studies.
- Finally, under no circumstances will the general reputation of a test or other selection procedures or casual reports of its validity be accepted in lieu of evidence of validity.

The minimum technical standards called for in the guidelines of all tests are that validity studies should be based on review of information about the job (a job analysis). The technical standards differ somewhat for the type of validation study. Tables 5.1, 5.2, and 5.3 summarize these standards by validation method.

**Table 5.1 EEOC Technical standards Guidelines for the criterion validation method**

Technical Standard for Criterion-Related Validation Studies	
1.	The study must be technically feasible. It must be possible to get an adequate sample size to provide a scientifically sound result. However, an employer is not required to hire or promote individuals in order to be able to conduct a criterion-related study.
2.	Whether the study is to be concurrent or predictive, the sample subjects should be representative of the individuals who might reasonably be expected to fill the positions being studied.
3.	In general, the guidelines indicate the finding of a significance level $P \leq 0.05$ to be acceptable.
4.	However, users should evaluate each selection procedure to assure that it is appropriate for operational use. In general, the greater the magnitude of the correlations found between the job behaviors and the tests, and the greater the number of job behaviors predicted by a particular test, the more appropriate it is for implementation. Selection procedures derived from studies with large sample sizes and low correlations, and sole reliance on a selection instrument that is related to only one of many critical job behaviors will be subject to close review.
5.	Users must avoid use of techniques that can lead to inflated validities for selection procedures. Examples include reliance on a few selection procedures or criteria when many were studied, and use of the statistics from one sample when they may not have held up well on cross-validation. The Guidelines recommend large samples and use of cross-validation.
6.	The Guidelines call for the maintenance of "fairness" in selection procedures. Essentially, unfairness results when members of one group characteristically obtain lower scores on a selection procedure than members of another group, but the differences in scores on the selection instrument are not manifest in differences in job performance. The guidelines call for investigation of the fairness of selection procedures whenever a selection device has adverse impact.

**Table 5.2 EEOC Technical standards Guidelines for the content validation method**

Technical Standard for Content Validation Studies	
1.	Consideration must be given to the appropriateness of content validity strategy. Such a Strategy is not appropriate when the job tasks represent knowledge, skills, and abilities that an employee is expected to learn on the job. It is also not appropriate for demonstrating the validity of selection procedures that claim to measure traits or constructs such as intelligence, aptitude, personality, common sense, judgment, and leadership.
2.	The job analysis must focus on the important work behaviors, their relative importance across all behaviors, and the products of such work behaviors. To be included in a work sample, the behaviors must be observable, and some aspect of them must be measurable. The work behaviors selected for measurement should be critical and/or important work behaviors that constitute most of the job.
3.	To demonstrate content validity of a selection procedure, it must be shown that the behaviors are a representative sample of behaviors of the job or that the selection procedure offers a representative sample of the work product of the job. For selection procedures measuring a skill or ability, the procedures must closely approximate an observable work behavior or work product. The closer the content and the context of the selection tests are to work samples and work behaviors, the more suitable they are for showing content validity.
4.	Whenever feasible, measurement of the reliability of the Selection procedures should be carried out.

**Table 5.3 EEOC Technical standards Guidelines for the construct validation method**

Technical Standard for Construct Validity Studies*	
1.	The Guidelines recognize that establishment of construct validity is a more complex strategy than either content or criterion-related validity, and that there was, at the time of Guidelines' publication, a lack of literature extending the concept to employment practices.
2.	Therefore, the job analysis must be carried out in a fashion that allows the identification of constructs underlying the important job behaviors. Each construct discovered should be named and defined to distinguish it from all other constructs so discovered.
3.	Selection procedures should then be developed or identified that measure the work behavior constructs. The users must then show that the selection procedures are related to the work behavior constructs and that the work behavior constructs are validly related to the performance of important or critical work behaviors.
4.	The Guidelines allow limited use of construct validity studies. "Until such time as professional literature provides more guidance on the use of construct validity in employment situations, the Federal agencies will accept a claim of construct validity without a criterion-related study ... only when the selection procedure has been used elsewhere in a situation in which a criterion-related study has been conducted and the use of a criterion-related validity study in this context meets the standards for transportability of criterion-related validity studies set forth above..."

\* see Figure 52

## Physiological Validation

The validation models identified in the EEOC Guidelines (EEOC, 1978) are based on the American Psychological Association standards for validating educational and psychological tests (A.P.A., 1985; A.P.A., 1987). A major difference when validating physical tests is the use of physiological, not psychological, tasks. Physiological tests differ from educational and psychological tests.

The goal of physiological validation is to match the worker with the physiological demands of the job. An essential element of this process is the quantification of the task's physiological stress. The recent court ruling of *Lanning v. SEPTA* (U.S. 3<sup>rd</sup> Circuit 1999) gives legal support to physiological validation. The case is discussed in greater detail in Chapter 7 of this State of-the-Art Report (SOAR). A key issue in the *Lanning v. SEPTA* case was setting a valid aerobic fitness cut-



score. The recommended cut-score represented a  $\text{VO}_2\text{max}$  of 42.5 ml/kg/min. The court ruled the standard to be unacceptable because the test developers failed to identify the minimum aerobic capacity demanded by the job.

The tradition and “standard practice” used to validate criterion-related physiological tests is to use the metric of the dependent variable (i.e., the criterion test) as the basis for evaluating a subject’s work capacity, which is sampled with the predictor test. The metric of the criterion variable has physiological significance. This physiological test validation methodology is clearly illustrated with body composition and  $\text{VO}_2\text{max}$  concurrent test validation research. To illustrate, in 1951, Brozek and Keys (1951) not only reported the concurrent validity coefficient between the predictor test, skinfold fat, and the criterion variable, hydrostatically measured percent body fat, but also published the first regression equation providing a valid model to interpret a subject’s skinfold fat measurement by the more meaningful metric of percent body fat. As another example, the maximum treadmill test following a standard protocol is a method of measuring  $\text{VO}_2\text{max}$ . These concurrent validation studies (Bruce, Kusumi, & Hosmer, 1973; Foster, Jackson, & Pollock, 1984; Pollock et al., 1976) published a regression equation with functions to estimate  $\text{VO}_2\text{max}$  (ml/kg/min) from treadmill time. The metric used to interpret aerobic fitness is  $\text{VO}_2\text{max}$ , not elapsed treadmill time. The next section of this chapter examines differences in the validation of physiological and psychological tests.

## Differences in Physiological and Psychological Test Validation

Although the psychological-based validation strategies outlined in the EEOC Guidelines are suitable for validating physical tests, there are at least three important differences. These include the test metric used, the work task definition, and the matching of the worker to the demands of the task.

**Test Metric**—The first major difference between psychological and physiological tests is the test’s metric. Typically, the metric of physiological tests is a ratio measurement scale. In contrast, scaling of psychological tests is either ordinal or interval. The units of measurement of physiological tests include percent body fat, oxygen uptake, caloric expenditure, force exerted, pounds lifted, weight load transported, and various types of power output, to name a few. The unit of measurement has physiological significance. In contrast, the unit of measurement of psychological tests is typically an individual’s response on a knowledge test or response to some type of scale (e.g., Lickert scale). The unit of measurement on psychological tests is of little importance. This is evidenced by the common practice of transforming scores on psychological tests from the original metric into some form of standard score with a known mean and standard deviation, such as 500 and 100. The person’s score is interpreted relative to the mean and standard deviation of the test. In contrast, a physiological test is not only interpreted with the mean and standard deviation of a population, but the value can also have an important physiological meaning. For example, a  $\text{VO}_2\text{max}$  of 20 ml/kg/min not only signifies a person has low fitness by normative standards but also indicates that the person lacks the physiological capacity to perform work tasks with an energy cost that exceeds the person’s low aerobic capacity,

**Accurate Quantification of Work Demands**—A characteristic of physiological test validation is that the physical demands of work tasks can often be objectively measured. This is because of the capacity to define the physical demands of the work task. Extensive physiological research has defined the energy expenditure of a host of occupational, recreational, and fitness tasks by measuring oxygen consumption while doing the tasks (Durnin & Passmore, 1967; Passmore & Durnin, 1955). These energy-cost tables are published in basic exercise physiology texts (Astrand & Rodahl, 1986; Brooks & Fahey, 1984; McArdle, Katch, & Katch, 1991; Wilmore & Costill, 1994). The forces required to “crack” valves and push or pull objects can be measured with torque wrenches and electronic load cells (Jackson, Osburn, Laughery, & Sekuls, 1998; Jackson, Osburn, Laughery, & Vaubel, 1992). The demands of materials-handling tasks can be defined by weight load, type of lift, lift rate, and distance transported (Jackson, Osburn, Laughery, & Young, 1993a; Waters et al., 1999; Waters, Putz-Anderson, Garg & Fine, 1993). These objective data define the physiological stress demanded by work tasks.

**Match the Worker to the Physiological Demands of the Task**—A final difference between physiological and psychological test validation is the capacity to match the worker to the physiological demands of the work task. Once the demands of the work task are known, the next step of a physiologically-based validation strategy is to determine if a worker has the capacity to meet the demands of the task. This was the method used to define the minimum energy cost (i.e.,  $\text{VO}_2\text{max}$ ) required for fire-fighting (Sothmann et al., 1990). This research showed individuals with a  $\text{VO}_2\text{max}$  below 33.5 ml/kg/min were unable meet the demands of firefighting. A goal of ergonomic research has been to define the strength levels needed to do industrial tasks safely (Keyserling et al., 1980; Keyserling, Herrin, & Chaffin, 1980). The next sections of this chapter discuss these methods in more detail.

## Physiological Validation—Test Fairness

The goal of a physiological criterion-related strategy is not only to estimate the validity of the test but also determine the minimum physiological level required by the task. A second important element of this approach is the physiological interpretation of the obtained data analyses. Interpretation of the statistical results of validation research with relevant physiological theory and published research provides a scientific rationale to explain the results. Failure to do this leaves the validation results open to question.

An important issue to resolve in a criterion-related study is whether the preemployment test is fair. Unfairness is defined as a situation in which members of a protected group obtain lower scores on a preemployment test than members of another group, but the difference in scores is not reflected in differences in the criterion of job performance (EEOC, 1978). This is called the Cleary test of fairness and is affirmed by showing that the regression line that defines the relationship between the preemployment test and the criterion is common to both groups. The statistical procedure is to test for homogeneity of regression slopes and intercepts (Arvey & Faley, 1988; Jackson, 1989; Pedhazur, 1997). The literature provides examples of the use of this test (Arnold, Rauschenberger, Soubel, & Guion, 1982; Reilly, Zedeck, & Tenopir, 1979).

Although the Cleary test may evaluate the fairness of an employment test, the analyses can also provide a physiological interpretation of the employment test. The Cleary test is the method of determining whether a common regression equation can be used to explain the relationship between the predictor and criterion tests of two groups. In physical test validation, the two groups are typically male and female applicants. The data analysis strategy is first to determine whether the two groups share a common regression slope and then decide whether the groups' regression intercepts are within chance variation. Multiple regression is the statistical model used to test for fairness. This multivariate analysis involves dummy-coding the group variable (e.g., female = 0, male = 1) and forming a group by predictor test interaction term (Pedhazur, 1997). The statistical strategy used is to generate a full multiple regression consisting of the three variables—

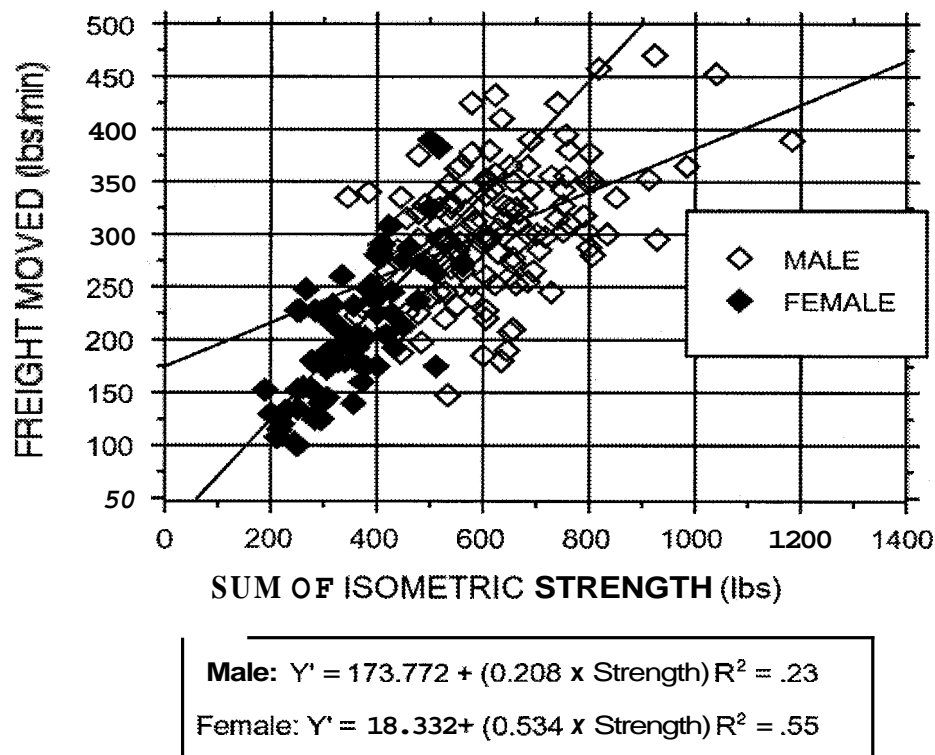
1. a predictor test
2. a dummy-coded group variable, and
3. an interaction term, which is the product of the group and test variables.

The next step is to generate two restricted regression models: the first with two independent variables, the group variable and the predictor test; and second, with just the predictor variable. The statistical test used to evaluate group differences in slopes and intercepts is to evaluate changes in  $R^2$  between the full and restricted models. Pedhazur (1997) outlines these statistical methods and tests of significance. These methods are illustrated next with physiological data. Also shown are the role and importance of the physiological interpretation of the results.

**Group Difference in Regression Slopes**—A task analysis of freight mover tasks showed that rapidly moving packages from a container to a conveyor belt was a physically demanding task (Jackson et al., 1993a). A work-sample test was developed to duplicate the demands of this repetitive transport task. The task involved moving packages that ranged in weight from about 15 to 80 pounds. The distribution of package weights was representative of the weight distribution encountered by workers. A work-sample test duplicated work demands of the task. Exercise heart rate was measured to ensure the work rate of the simulation test was representative of the actual work rate. The subjects were instructed to work at a brisk rate consistent with their fitness and not to move packages that exceeded their capacity.

Figure 5.3 is the bivariate relationship between the predictor test (sum of isometric strength) and the criterion test (materials transport, expressed in a metric of power output, the pounds of freight transported per minute). The data are contrasted by gender. Analysis of these data showed that male and female regression lines were not parallel. The  $R^2$  change between the full model and restricted model of the strength test and dummy-coded gender variable was 0.04, which was statistically significant ( $F_{(1, 199)} = 18.96$   $p < 0.01$ ). The graph shows that the slope for the female subjects (0.534) is more than twice as steep as the slope for male subjects (0.208).

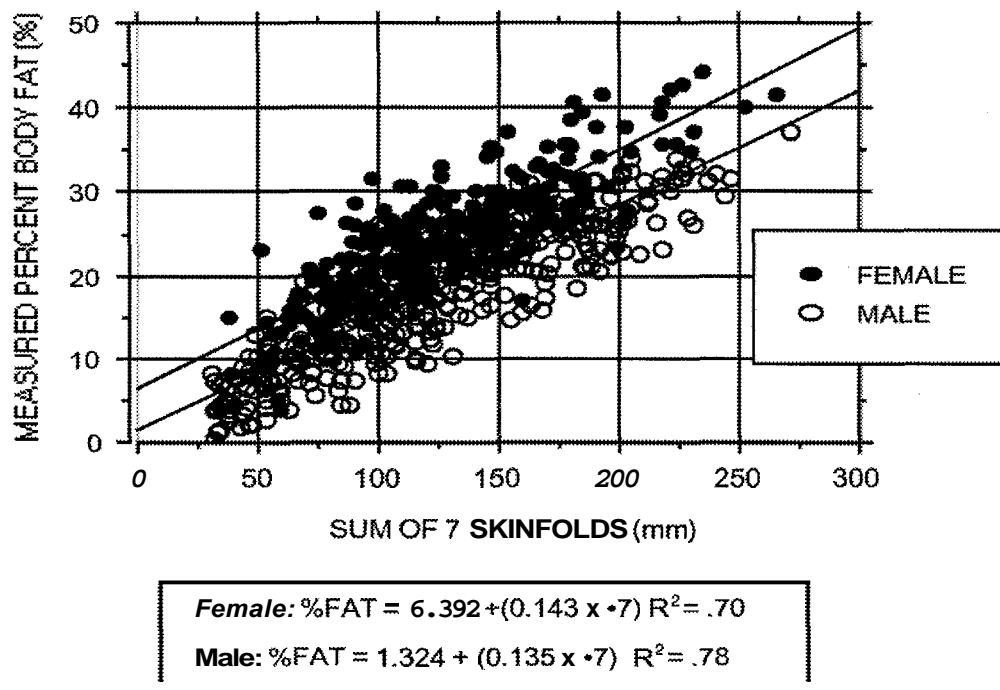
A strict interpretation of the Cleary test would indicate that the strength test was unfair, but a physiological interpretation of the data gives a clearer view. Post hoc examination of the data showed that many females could not lift and transport the heavier packages. The lift weight exceeded their strength capacity. The steeper female slope showed that individual differences in strength were more important for females than males. The stronger women could lift the heaviest weight



**Figure 5.3** Test for fairness, example of significant differences in male and female regression slopes

loads while the weaker women could not. A major determinant of the female capacity to move freight was the subject's strength-dependent capacity to lift heavy loads. In contrast, most men had the physiological capacity to lift and transport the heaviest loads. These physiological data would be important information for setting a cut-score consistent with the demands of the task. The data could also have important ergonomic implications that could lead to job redesign, such as a company policy limiting the weight of packages they would transport.

**Intercept Differences**—The second part of the Cleary test is to evaluate differences in regression intercepts. Figure 5.4 shows a physiological example of intercept differences in the form of the scatterplot of published male and female body composition data (Jackson & Pollock, 1978; Jackson, Pollock, & Ward, 1980). The independent variable is the sum of seven skinfold measurements, and the dependent variable is percent body fat measured by the underwater weighing method. The figure shows that the slopes of the male and female regression lines are parallel; the differences in slope are within random variation ( $F_{(1,675)} = 1.25$ ;  $p > 0.05$ ). The  $R^2$  difference between the full model and restricted model with gender and the sum of skinfolds was 0.0004. Adding the dummy-coded gender variable to the sum of skinfolds accounted for more than 12 percent of percent fat variance ( $F_{(1,675)} = 398.75$ ;  $p < 0.01$ ). As these data show, the significant intercept difference indicates that for a given score on the predictor test (sum of skinfold fat), the criterion score of one group can be expected to be systematically higher, which in this instance is measured percent body fat. The regression lines differed by an average percent body fat of about 6 percent.

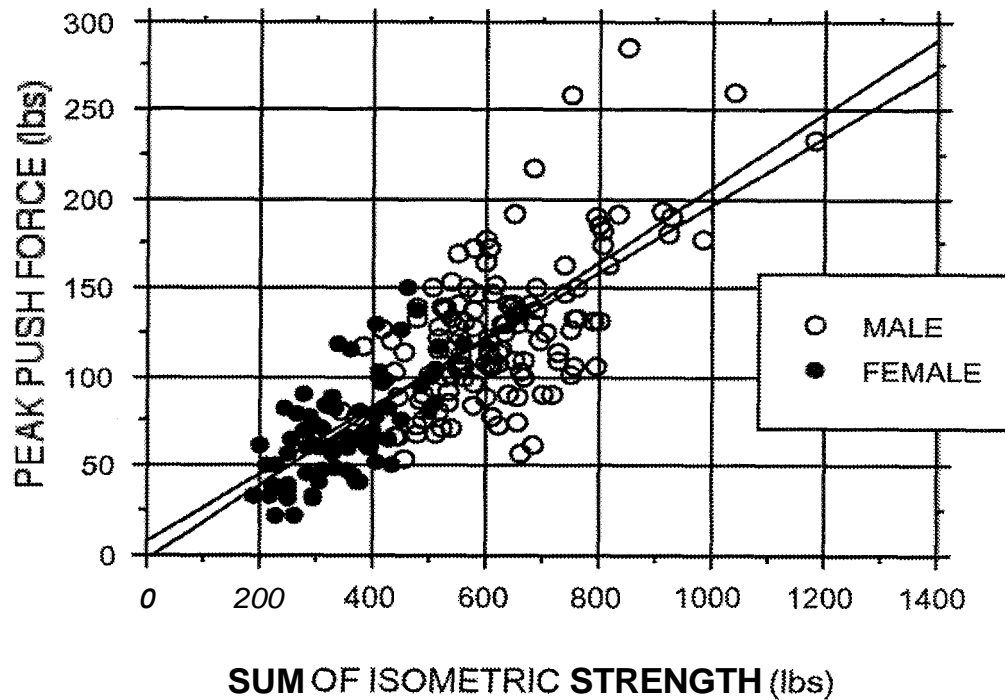


**Figure 5.4** Test for fairness, example of parallel regression slopes, but significant differences in male and female regression intercepts

A “blind” application of the Cleary test would indicate that the test was unfair. A physiological interpretation of these results provides a clear rationale for the intercept difference. Skinfold fat measures subcutaneous fat, but the body has two types of fat, subcutaneous and essential fat. Hydrostatically determined percent body fat measures both sources of body fat. It is well established that the essential fat of women is greater by about 7 percent of body mass than that of men (McArdle et al., 1996). The physiological explanation for the gender difference in intercepts can be explained by differences in essential fat.

Although this body composition example does not represent a work-sample test, the use of body composition tests has been an interest of Military researchers (Marriott, 1992). It is well-documented that percent fat is inversely related with strenuous tasks that involve moving the body. This body composition example shows that if percent body fat is used to evaluate male and female performance on common physical tasks (e.g., running, climbing), the test must be expressed in the physiological metric of percent body fat, not the sum of skinfold fat. In contrast, if the goal is to evaluate fitness rather than the capacity to meet the demands of a work task, gender-based standards are appropriate (Gettman, 1993).

**Common Slope and Intercept**—The example provided in this section illustrates the homogeneity of male and female regression lines for the predictor and criterion tests. Figure 5.5 gives the scatter plot of the male and female relationship between isometric strength and peak push force. A task analysis showed that push force was a physically demanding task required of workers who moved freight containers (Jackson et al., 1993a). The mean push force of the males was 124.6 (SD = 42.2)



*Figure 5.5 Test for fairness, example of homogeneity of male and female regression slopes and intercepts*

compared with a mean of 70.0 (SD = 28.3) for the females. This difference was statistically significant ( $F_{(1,205)} = 99.89; p < 0.01$ ). The figure shows that the male and female regression lines are similar. Statistical analysis showed the slopes ( $F_{(1,205)} = 1.50; p > 0.05$ ) and intercepts ( $F_{(1,203)} = 2.00; p > 0.05$ ) of the male and female regression lines were not statistically significant. The group and group-by-strength variables accounted for less than 0.1 percent of the push-force variable. This demonstrated that differences in the regression lines shown in the figure were random variance. This analysis demonstrated that a single regression line can be used to estimate push force from isometric strength, and documented that the gender mean difference in work task performance depended on strength, not gender.

## Physiological Validation—Cut-Score

Once the predictor test has been shown to be valid, the next step of a physiological validation strategy is to define performance on the predictor test associated with the desired level of performance on the criterion. **An** important and often difficult part of this analysis is defining the critical level of performance on the criterion variable. In some instances, a clear definition of an essential task is apparent, for example, lifting a 75-pound industrial valve from the ground to the back of a truck. In other instances, the physiological demands of a task can be difficult to quantify accurately. Shoveling coal is a physically demanding task of coal miners (Jackson & Osburn, 1983), but what level of intensity and duration of shoveling are suitable? Firefighter work simulation tests are

timed tests that involve completing several firefighter tasks. Although a firefighter test may be clearly content valid, a more difficult phase of the validation process is to determine the time that signifies successful fire-fighting capacity (Jeanneret & Associates, 1999).

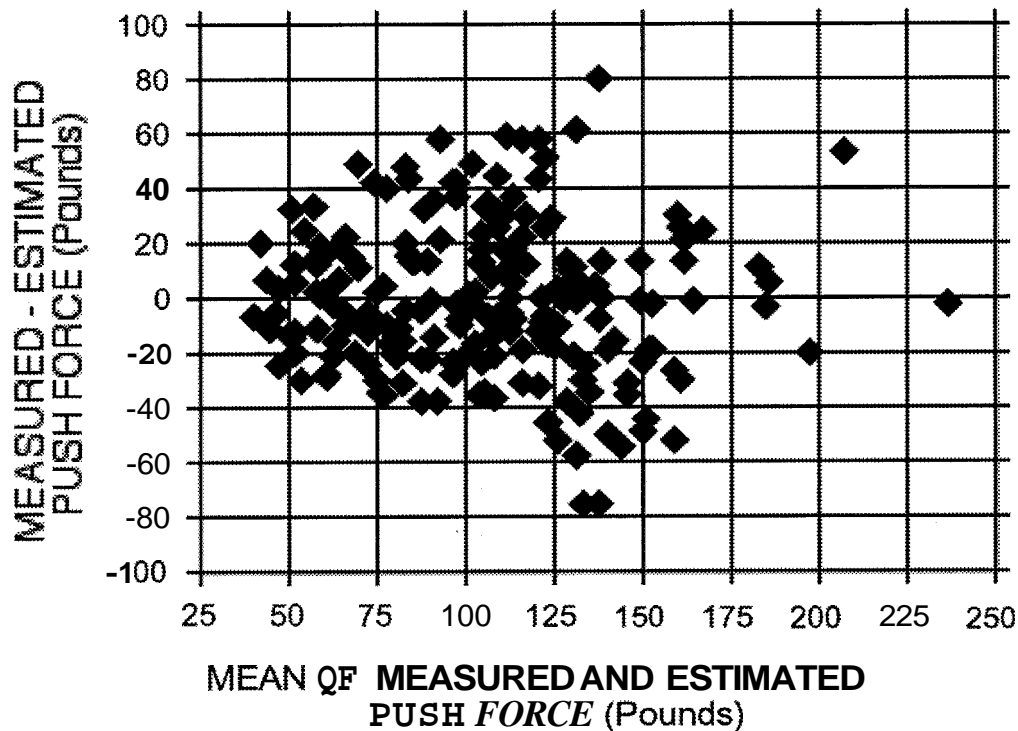
Regression models provide valid statistical methods of estimating physiological capacity, within a defined degree of accuracy, from a predictor test or combination of tests. Simple linear and nonlinear regression models are used with a single predictor test, and multiple regression models are used with several predictor tests (Pedhazur, 1997). This is a well established physiological test validation method (ACSM, 1991; Astrand & Ryhming, 1954; Brozek & Keys, 1951; Bruce et al., 1973; Durnin & Wormsley, 1974; Foster et al., 1984; Jackson, 1990; Jackson & Pollock, 1978; Jackson et al., 1980; Pollock et al., 1976). The following provides regression examples of defining physiologically based standards with continuously scaled and pass/fail criterion variables.

**Continuously Scaled Criterion**—This first example shows the use of simple linear regression to define the strength needed to generate the push force required by a task. The job analysis (Jackson et al., 1993a) showed that one physically demanding job of freight workers was pushing or pulling containers loaded with freight. As part of the job analysis, an electronic load cell defined the peak force required to move freight containers that varied in weight. The subject's peak push force was measured with an isometric push test that simulated the position used to push containers. Figure 5.5 shows the scattergrams with the male and female regression lines. As shown earlier, the difference between the slopes and intercepts of the male and female regression lines were within chance variation which supports the fairness of using a single regression line to define this relationship. The regression equation is—

$$\begin{aligned} &\textbf{Push Force Regression Equation } (R=0.78, SEE = 29.0 \text{ lbs}) & (1) \\ &\textit{Push Force (lbs)} = 2.031 + (0.198 \times \textit{Strength}) \end{aligned}$$

The regression equation provides a valid model for defining the strength needed to generate the push force needed to move containers of the criterion weight. Once this is known, the strength associated with this push force can be determined. To illustrate, assume the criterion push force was defined to be 100 pounds of force. The regression equation shows that a strength score of 495 estimates a push force of 100 pounds.

The goal of a physiological model of validation is to define the minimum physiological capacity demanded by the work task. The regression model provides empirical evidence to define a physiologically defined cut-score within a defined level of probability. Although physiological tests scores typically yield higher criterion-related validity coefficients than psychological tests, they still have substantial prediction errors. Figure 5.6 shows the predictor errors associated with the push force task. Provided is an Altman-Bland plot (Altman & Bland, 1983; Altman & Bland, 1986) of the push force data estimated from isometric strength (see Figure 5.5). The Altman-Bland method plots the difference between the residual scores ( $\bar{Y} - Y'$  which is measured estimated push force) by the average of measured and estimated push force. Although the correlation between the criterion, push force, and predictor, isometric strength, was high, 0.78, the Altman-Bland plot shows that defining the physiological criterion is not error free. The variability on the Y axis is defined by the standard error of estimate of the regression analysis, which, in this example, is 29 pounds of push force.

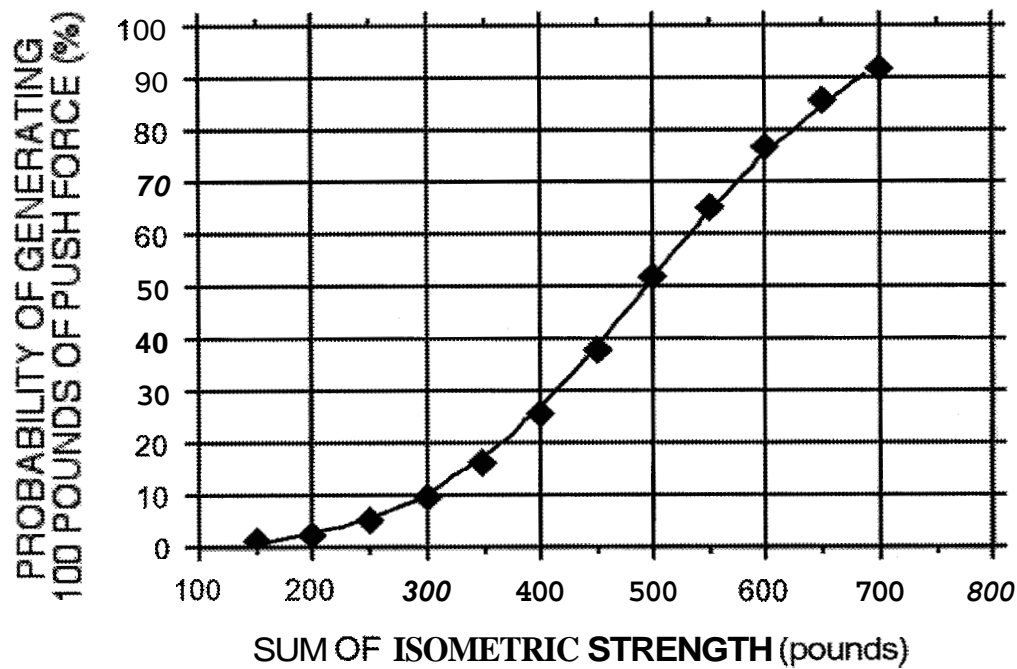


**Figure 5.6** Reprinted, by permission, from Altman, D. G., Blandm J. M. (Altman-Bland plot of prediction residuals (measured – estimated) contrasted by the average of measured and estimated maximum push force). pp. 307–310, © by the Lancet Ltd., 1986.

Because the correlation between a predictor variable and the criterion test is always less than 1, there will always be prediction errors. The standard error of estimate provides an estimate of the variation in prediction error. Although it is not possible to define an exact physiologically-based cut-score, it is possible to define a standard with a defined degree of probability. The regression equation (Equation 1) provides a valid model that defines the relationship of strength with push force. As shown earlier, 495 pounds is associated with a push force of 100 pounds. Because the correlation between the two tests is less than perfect and there are prediction errors, only 50 percent of subjects with 495 pounds of strength would be expected to have the capacity to generate 100 pounds of push force. The regression model's standard error of estimate can be used to define the probability that someone, with a given level of strength, would meet the physiologically-based standard. Figure 5.7 shows the relationship between level of isometric strength and probability<sup>1</sup> of being able to generate 100 pounds of push force. The probability estimates provide additional data that can be used to define a physiological criterion that is congruent with the criticality of the task, and the mission and unique organizational characteristics.

**Pass-Fail Model** — Often, the criterion of job performance is scaled as a dichotomous variable. For example, manual lifting tasks are scored pass or fail—the applicant could or could not lift a given weight load (Jackson, Osburn, Loughery & Sekula, 1998; Jackson et al., 1992). Other examples are





**Figure 5.7** *Probability of being able to generate 100 pounds for push force for levels of strength*

endurance tasks at a constant power output. A manufacturing work task may require a worker to repetitively lift and transport weight loads at a given work rate governed by production speed. Individuals without sufficient physiological capacity would not be able to maintain the set pace. A task documented that refinery workers must close industrial valves during emergencies (Jackson, 1987; Jackson et al., 1992; Osburn, 1977). For some individuals, the task exceeded their physiological capacity and they fatigue quickly. For others, the task was within their physiological capacity. These fit individuals could continue work for extended periods of time. Demanding repetitive tasks at a set power output tend to produce a bimodal distribution—those who have and those who do not have the physiological capacity. This is illustrated in the literature (Jackson et al., 1992).

Logistic regression analysis (Hosmer & Lemeshow, 1989; Pedhazur, 1997) provides a model to physiologically validate tests when the criterion is a dichotomous variable. Logistic regression, like multiple regression, can use a single independent variable or several independent variables. A logistic regression model estimates the probability of group membership (e.g., criterion variable of pass or fail) given a score or scores on the predictor variable (Pedhazur, 1997). A public health landmark multiple logistic regression validation study was with the Framingham heart study (Kannel, McGee, & Gordon, 1976). The research objective was to identify and quantify cardiovascular disease risk factors. The logistic analysis not only established that cholesterol, blood pressure, glucose intolerance, and smoking were independent cardiovascular disease (CVD) risk factors, the statistical analysis also produced an equation with a function of estimating the probability of CVD risk for combinations of risk factors. Logistic regression analysis, like regression models with continuous variables, establishes the validity of the independent variable(s) and provides an empirical

model for defining the probability of group membership. The application of simple logistic regression analysis is illustrated below with a lifting task.

A task analysis of an oil production plant showed that lifting heavy valves from the floor to knuckle height was an important, physically demanding work task (Jackson, 1998). A work-sample test was developed to simulate the task. The work-sample test involved lifting several loads that varied in weight. The physical dimensions of the lift duplicated the work task. The test was scored pass or fail depending on the subject's ability to complete the lift. The predictor test was the sum of four isometric strength tests, arm, shoulder, torso, and leg strength. The goal of this physiological validation was to define the level of strength required for the lift task.

This validation method is illustrated with three weight loads, 60-, 90-, and 120-pound lifts. These weights represent industrial lifts ranging from moderately heavy to very difficult. The first step in this analysis was to determine whether lift success depended on strength. Table 5.4 provides the means, standard deviations, and sample sizes of the subjects who passed and failed the lift. Analysis of variance showed that lift success depended on strength and documented three, expected trends. First, the number of individuals who could lift the load decreased with the weight load. Next, the Analysis of Variance (ANOVA) documented that lift success for all three weights depended on isometric strength. The means for those who lifted the weight were significantly higher than for those who could not. Third, the mean strength of those who completed the lift increased with the weight load. These trends are consistent with physiological expectations.

**Table 5.4 Sample sizes, strength means and standard deviations, and analysis of strength differences of those who could and could not lift the weight**

Lift Weight	Lifted Weight		Did Not Lift Weight		ANOVA F-ratio
	N	M ± SD	N	M ± SD	
60-Pound	120	518 ± 197	16	196 ± 66	41.92'
90-Pound	93	579 ± 175	43	233 ± 101	118.89'
120-Pound	71	644 ± 141	65	301 ± 108	250.69*

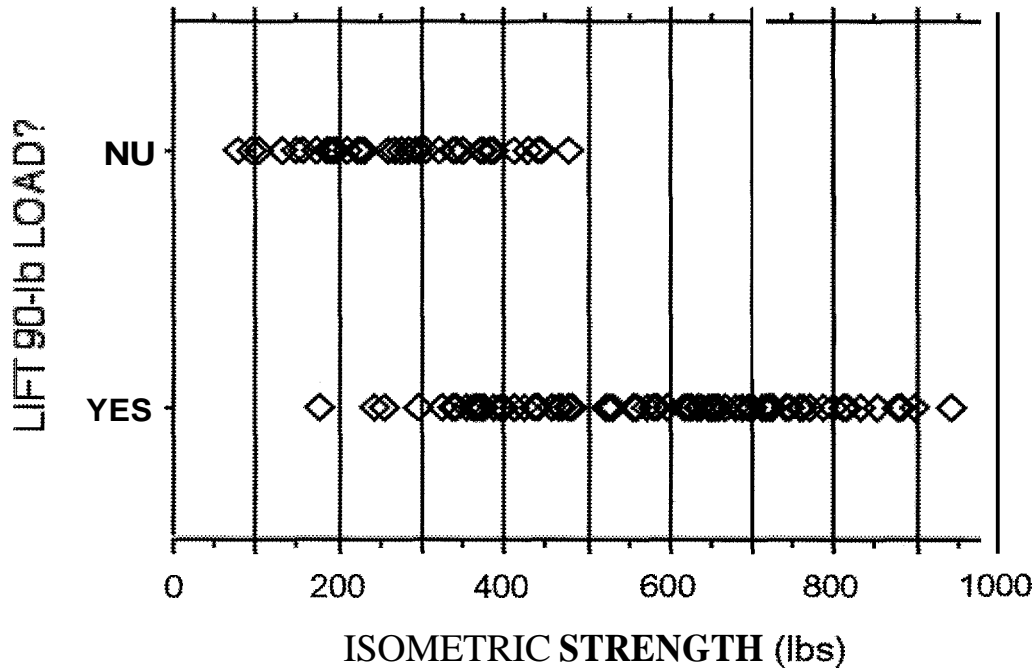
Figure 5.8 provides a scatter plot of the subjects' strength data contrasted with their 90-pound lift success. This plot shows the group difference in strength documented by the ANOVA but also shows an overlap in the strength of those who passed and failed the lift. Logistic regression analysis provides a model for estimating the probability of success on the criterion variable (i.e., lifting the load) for given levels on the predictor test (i.e., strength) or, in this example, the probability of being able to lift the load for a level of strength. The logistic regression analysis, which agreed with the ANOVAs (Table 5.1), showed that the regression weight for strength was significantly related to the probability of lifting the given weight. The equations for the three lift loads are—

$$\text{60-pound lift} \quad (2)$$

$$\text{Logit}(P) = (0.020 \times \text{Strength}) - 3.926$$

$$\begin{aligned} & \text{90-pound lift} & (3) \\ & \text{Logit}(P) = (0.017 \times \text{Strength}) - 5.689 \end{aligned}$$

$$\begin{aligned} & \text{120-pound lift} & (4) \\ & \text{Logit}(P) = (0.023 \times \text{Strength}) - 10.334 \end{aligned}$$



**Figure 5.8 Scatterplot of strength test of subjects who could or could not complete a 90-pound lift from floor to knuckle height**

Once the logistic equation is defined, Equation 5 estimates the probability of success (Pedhauzur, 1997). The term  $e$  in Equation 5 is the base of the natural logarithm; a value of  $Y$  2.718. Figure 5.9 graphically shows the probability of success in completing the lift for strength levels.

$$\begin{aligned} & \text{Logistic Probability Calculation Model} & (5) \\ & P = \left( \frac{e^{a+bX}}{1+e^{a+bX}} \right) \times 100 \end{aligned}$$

The logistic probability curves clearly show, as would be physiologically expected, that the strength needed to lift the load increases as the lift gets heavier. There is a 50 percent probability, for example, that someone with 200 pounds of strength could lift a 60-pound load. In contrast, only 10 percent of the subjects with 200 pounds of strength would be expected to lift 90 pounds. The likelihood of someone with 200 pounds of strength lifting 120 pounds is 0. The physiological levels needed to be 50 percent confident of lifting the 90- and 120-pound loads are about 350 and 450 pounds of strength, respectively.

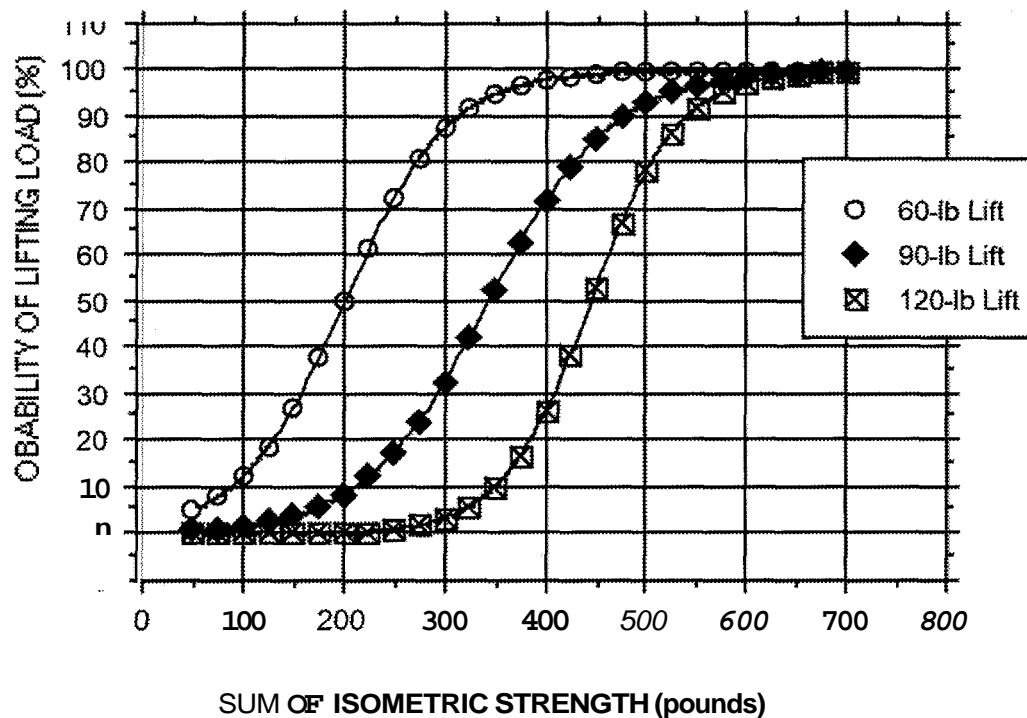


Figure 5.9 Logistic curves of the probability of being able to lift the weight load as a function of lifter strength

## Physiological Validation—Matching the Worker to the Job

The goal of physiological test validation is to select workers with the capacity to meet the demands of the job. This is consistent with ergonomic objectives designed to reduce the risk of job-related injuries (Ayoub, 1982). As has been shown in this chapter, the statistical models used to define the physiological stress of the task are less than absolute. This permits latitude in formulating physical cut-scores ranging from lenient to rigorous. The regression statistics, equations, and standard errors provide an empirical base for making the decision.

Although the regression models previously discussed can help define the degree of physiological stress, the difficult task of establishing a suitable cut-score for a criterion remains. The types of job performance criteria listed in the Uniform Guidelines that may be suitable are supervisory ratings, production rate, error rate, tardiness, absenteeism, and success in training. According to the Guidelines, this is not an inclusive list of criteria. Other examples of criteria used to validate physical tests include accidents (Reilly et al., 1979), field performance (Reilly et al., 1979), injury rates (Gilliam & Lund, 2000; Keyserling et al., 1980; Keyserling et al., 1980); lost time due to sickness or injury (Rayson et al., 2000a; Rayson et al., 2000b); and job-related work tasks (Arnold et al., 1982; Jackson, Osburn, & Laughery, 1998; Jackson, Osburn, & Laughery, 1984; Jackson et al., 1992; Jackson, Osburn, & Laughery, 1991; Jackson, Zhang, Laughery, Osburn, & Young, 1993b; Rayson, 2000a; Rayson, 2000b).

A crucial element of any evaluation strategy is the selection rate of a protected group, which, in physical testing, is females. The physiological validation method supplements the process of defining an appropriate cut-score approach with scientific evidence. This validation approach seeks to find the minimum physiological level demanded by the task. The Uniform Guidelines (EEOC, 1978) allow the use of rational judgment in setting a valid cut-score. An objective of the physiological validation process is to provide a scientific explanation of the validation results. Included in this process is the establishment of a sound cut-score. Receiver operator characteristic (ROC) analysis (Hulley, 1988) is one method used to establish physiological cut-scores. It supplements the regression results by defining a cut-score consistent with a strategy of maximizing either test sensitivity or specificity.

A ROC is a graphic analysis used to establish a trade-off between test sensitivity and specificity. If the goal is to maximize test sensitivity, the proportion of true positives (i.e., those who can meet the physiological demands of the work), the ROC would be a plot of test sensitivity by  $1 - \text{specificity}$ , which is the proportion of false positives. False positives are those identified by the test with the physiological capacity to meet the demands of the task but who cannot meet the demands. In this context, the ROC curve provides a rational method of selecting a cut-score based on a balance between high sensitivity and low specificity. The interested reader is directed to another source (Wellens et al., 1996) for the application of ROC analysis for establishing a physiological cut-score. The objective of that study was to find the body mass index (ratio of weight and height) that defined the obesity levels of 25 percent and 33 percent body fat content, determined hydrostatically, for men and women, respectively.

Several factors are considered when establishing physiologically based cut-scores. The following is a nonexhaustive list of conditions that may determine whether a lenient or rigorous cut-score is selected—

- **Adverse Impact**—The first concern is adverse impact. Consideration must be given to the number of the protected group that the standard screens out.
- **Risk of Injury**—Subjecting workers to physical demands increases the risk for work-related injuries. Numerous studies (Cady, Bishoff, O'Connell, Thomas, & Allan, 1979; Gilliam & Lund, 2000; Herrin, 1986; Keyserling et al., 1980; Liles et al., 1984; Snook, Campanelli, & Hart, 1978; Snook & Ciriello, 1991) show that the risk of musculoskeletal injury increases as the demands of the task approach the worker's maximum physiological capacity.
- **Physiological Interpretation of the Validation Results**—An important element of a physical test validation study is to establish the congruence among the validation results, published research, and physiological theory. It is critical to provide a sound physiological explanation of the validation results. Failure to be able to interpret the results by accepted academic standards leaves the decision open to question.
- **Environmental Conditions**—Often, the location at which the validation study is conducted will be different from the work environment. For example, firefighter tests are not administered in burning buildings, the source of demanding work. Environmental conditions (e.g., heat) that increase the demands of the task justify more rigorous standards.
- **Workforce Numbers**—The number of workers available at the work site can affect the rigor of a cut-score. A more lenient standard might be considered when several workers are available to do the work. Although a lenient selection standard would increase the probability that a worker cannot meet the most physical demands of the job (i.e., a false positive), it may

not be a serious problem if others are available to do the work. The stronger workers can help with the most demanding tasks. In contrast, a more rigorous standard might be considered if a worker does not have help.

- **Criticality of the Job**—In some jobs, the failure to meet the demands of a job can be dangerous. The dummy drag test is a common item of a preemployment firefighter test. This is a critical task because the inability to perform it successfully can be life threatening.
- **Workforce Productivity**—Selecting workers with a higher physiological capacity can increase an organization's productivity. The data in Figure 5.3 show that the amount of freight a worker was capable of moving was related to the worker's strength capacity. This was one of the factors considered by a freight company to initiate a preemployment test program?

---

## Published Validation Studies

Although many preemployment tests have been completed, most are not in the published literature. The completed validation study often is a technical report to the governmental agency or private company that funded the project, and many organizations consider these privileged. Hogan (1991b) provides an extensive list of these unpublished reports. The following sections summarize the published validation research.<sup>3</sup>

### Outside Craft Jobs

One of the first published concurrent validation studies was for outdoor telephone craft jobs that involved pole-climbing tasks (Bernauer & Bonanno, 1975; Reilly et al., 1979). The issues leading to the development of this study were the large differences between male and female workers in turnover and accident rates. After 6 months, 43 percent of the women left the outdoor craft jobs compared with only 8 percent of the males. More important, women sustained substantially more injuries than men from falls while climbing or working on poles.

An extensive job analysis showed that pole climbing was an essential, physically demanding work task. Bernauer and Bonanno (1975) evaluated the factor composition of 40 tests and anthropometric measures on a sample of 241 job applicants. They developed a six-item battery consisting of reaction time, grip strength, percent body fat, step test performance, balance, and sit-ups. They found that the balance and step tests significantly differentiated successful from unsuccessful students enrolled in pole-climbing school.

Reilly and associates (Reilly et al., 1979) extended this work by completing two concurrent validation studies. In the first experiment, several anthropometric and physical performance tests were administered to 83 male and 45 female candidates for outdoor telephone craft jobs. Two validation criteria were used in this experiment. The first, general task performance, was the average of two supervisor performance ratings of the candidate's performance during the 5-day pole-climbing school. Job analysis data were used to construct the rating scale. The second criterion was a dichotomy of those who were on the job 6 months after placement and those who were not. Using

the criterion of general task performance, stepwise multiple regression isolated a three-predictor battery consisting of dynamic arm strength, reaction time, and Harvard bench step time. The analysis yielded a multiple correlation of 0.45. The statistically significant zero-order correlations between the job tenure criterion and these tests were dynamic arm strength, 0.36; reaction time, 0.19; and bench step time, 0.18. Further analysis showed that a common regression line defined male and female performance that met the important criteria of job fairness.

The second experiment used a larger sample of employees who represented the whole company. The criterion of pole-climbing training success was changed to be consistent with changes introduced in the pole-climbing course. The second study included four different criterion measures of job performance—

1. time to complete the pole-climbing school,
2. completion of pole-climbing school (a number withdrew from the course),
3. field observations of pole-climbing proficiency, and
4. accidents for 6 months after entering outdoor craft work.

The second sample consisted of 78 female and 132 male pole-climbing school applicants.

Multiple regression selected a three-item battery consisting of body density estimated from skinfold fat, balance, and an isometric arm strength test. The criterion was time to complete the course. The significant correlations among the three tests and the four criteria were time to complete the course, 0.46; training dropout, 0.38; field observations for the female sample, 0.53; and accidents, 0.15. Further analysis showed that the same regression equation was equally valid for both males and females.

## Firefighters

Nearly all major fire departments have a physical ability preemployment test (Landy & Investigator, 1992). Considine and associates (Considine et al., 1976) published the first physical test battery for screening firefighter applicants. The test battery evolved from an occupational task analysis that surveyed, rated, and analyzed 81 tasks performed by firefighters. The authors selected a construct validation strategy. The constructs identified through the task analysis were dynamic strength, static strength, agility, total body coordination, cardiorespiratory endurance, muscular endurance, eye-hand coordination, and total body speed.

The sample of the first study consisted of 191 males who were tested on body composition measures, general physical performance tests, and eight job sample tests. A factor analysis of these data produced three general factors. The factor names and tests representing each factor were factor 1, the ability to handle the body weight measured by percent body fat, obstacle run, and flexed-arm hang; factor 2, muscle power measured by the hose lift, man-lift-and-carry, and stair climb work sample tests; and factor 3, body structure measured by fat-free weight and height.

A major purpose of the second study was to analyze the test battery for racial bias. Based on the results of the first study, nine tests were administered to 165 firefighters and 19 candidates. Data analysis showed that African-American and white subjects did not differ on any of the tests. These

data were factor analyzed producing three common factors. The final recommended battery consisted of four work sample tests, and one fitness test; the flexed-arm hang. The work sample tests were modified man-lift-and-carry that simulated rescuing a trapped victim; stair climb that simulated climbing the stairs in a building; obstacle run that simulated moving the body through confined spaces; and hose couple that involved coupling three hoses to a hose couple.

Davis and associates (Davis, Dotson, & SantaMaria, 1982) examined the relationship between simulated firefighting tasks and physical performance measures. The sample consisted of 100 randomly selected men from the population of Washington, DC, firefighters. The physical performance measures included body composition, general fitness, aerobic fitness, and cardiovascular variables. The five work-sample tests came from the job analysis of firefighter work tasks and involved handling a ladder, lifting and transporting a 33.1-kilogram load up five flights of stairs, pulling a 23.5-kilogram hose roll from the ground up to and through the fifth-floor window, carrying and dragging a 53-kilogram dummy down five flights of stairs, and using a sledge hammer to simulate forceful entry.

Canonical correlation showed that two, independent dimensions defined the relationship between the physical performance variables and firefighter work-sample tests. The first canonical dimension ( $R_c = 0.79$ ) represented a physical work capacity factor that reflected the muscular strength and endurance, and maximal aerobic capacity elements of the simulated work-sample tests. The second dimension ( $R_c = 0.63$ ) represented a resistance to fatigue factor and the ability to complete the work tasks quickly. Multiple regression selected two physical performance batteries (laboratory and field batteries) to estimate each work-sample dimension. The field test battery for the physical work capacity factor consisted of push-ups, sit-ups, and grip strength. The validity of the field battery ( $R = 0.73$ ) was lower than the five-item laboratory battery ( $R = 0.95$ ) that added submaximal oxygen pulse and maximum heart rate to the battery. The three-item field test of the second factor included estimated percent body fat, lean body weight, and  $\text{VO}_2\text{max}$  estimated with a step test ( $R = 0.77$ ). The laboratory test added maximum heart rate and treadmill performance and increased the validity ( $R = 0.89$ ) of the resistance to fatigue work sample factor.

The physiological response of fire fighting has been the focus of many investigators. Exercise heart rate responses elicited by simulated and actual firefighting tasks confirmed that these tasks have a significant cardiovascular effect (Barnard & Duncan, 1975; Davis & Convertino, 1975; Lemon & Hermiston, 1977; Manning & Griggs, 1983; O'Connell, Thomas, Caddy, & Karwasky, 1986; Sothmann, Saupe, Jasenor, & Blaney, 1992). In a study during actual fire-suppression emergencies, Sothmann and associates (Sothmann et al., 1992) measured exercise heart rate and oxygen uptake on 10 male fire fighters. Their data showed that firefighters worked at an average of 88 percent ( $\pm 6\%$ ) of their measured maximum heart rate for an average duration of  $15 (\pm 7)$  minutes. The average energy cost of the firefighter emergency work task was a  $\text{VO}_2$  of  $25.6 \pm 8.7$  ml/kg/min, representing an intensity of 63 percent ( $\pm 14\%$ ) of  $\text{VO}_2\text{max}$ .

Sothmann and associates (Sothmann et al., 1990) examined the relationship between  $\text{VO}_2\text{max}$  and firefighting work tasks. A seven-item, content-valid fire suppression test was administered to 20 experienced fire fighters. The average energy cost of the firefighter simulation tests was  $30.5 (\pm 5.6)$  ml/kg/min. The work simulation required the firefighters to work at an intensity of 76 percent ( $\pm 8$ ) of  $\text{VO}_2\text{max}$ . The correlation between the elapsed time required to complete the firefighter work simulation test and measured  $\text{VO}_2\text{max}$  was  $-0.55$ . In a cross-validation study with 32 different male firefighters, successful work simulation performance depended on  $\text{VO}_2\text{max}$ . Of the 32



tested, seven firefighters could not complete the work sample tests. The  $\text{VO}_{2\text{max}}$  of five of the seven was below 33.5 ml/kg/min.

## Highway Patrol Officers

With an increasing number of women seeking employment as highway patrol officers, the objective of the study published by Wilmore and Davis (1979) was to find the minimum physical qualifications and develop a job-related preemployment test. They administered three different batteries of tests to 140 male and 16 female patrol officers. The laboratory and field test batteries included strength, flexibility, body composition, and cardiorespiratory endurance items. The job sample tests included a barrier surmount and arrest simulation, and a dummy drag that simulated dragging an injured victim 50 feet to safety.

The major differences between the field and laboratory batteries were that the 1.5 mile run replaced the maximum treadmill test, and body fat was estimated from skinfolds rather than measured by hydrostatic weighing. The laboratory test battery was significantly correlated with the dummy drag ( $R=0.66$ ) and barrier surmount and arrest simulation tests ( $R=0.68$ ). Replacing the laboratory tests with the field tests resulted in slightly lower correlations, 0.57 for the dummy drag, and 0.62 for the barrier surmount and arrest simulation tests. Although the fitness tests estimated work simulation test performance, test performance was not related to job performance consisting of supervisor ratings on 16 critical job tasks.

The data analysis showed that the officers were similar to the normal population in strength, body fat, flexibility, and cardiorespiratory endurance. An important result of the study was that the predominantly sedentary nature of the officer's job led to a rapid deterioration in physical fitness following his or her academic training, suggesting the need for an in-service physical conditioning program.

## Steel Workers

Arnold and associates (Arnold et al., 1982) developed a preemployment test for selecting entry-level steel workers. The task analysis documented that entry-level steel workers must do several different physically demanding tasks. The investigators used a combination of content- and construct-validation strategies. The job analysis identified the physically demanding work tasks required of the entry-level workers and categorized them by Fleishman's constructs of static strength, dynamic strength, and endurance (Fleishman, 1964). The selected candidate physical performance tests were those that theoretically measured these constructs.

The objective of the study was to determine whether the physical performance tests were related to the work-sample tests developed from the job analysis. The sample included 168 men and 81 women who were in their first 6 months of employment at three different plant locations. The job analysis showed that work tasks differed somewhat across the 3 sites, resulting in 11 work sample tests at 1 site and 12 at the other 2 sites. The average work-sample test performance was the criterion of work performance. In addition to the work-sample tests, each subject completed 10 physical performance tests sampling strength, flexibility, agility, balance, and cardiorespiratory endurance dimensions.

Multiple regression selected the physical performance tests most highly correlated with the work-sample criterion. For all three work sites, arm dynamometer strength was the most important predictor of work-sample test performance. The zero-order correlations between arm strength and work-sample test performance were consistently high—0.82, 0.85, and 0.85 for the three sites. Adding two more tests to the multiple regression models added little to the validity; the multiple correlations for the three predictor models increased to 0.87, 0.88, and 0.89.

The authors completed a utility analysis for the single arm strength test (Hunter, Schmidt, & Hunter, 1979). This analysis involved estimating the money the company would save by hiring workers who could do the work. Utility estimates were based on test validity and the monetary value was related to the variability of work performance. Using 1982 wage standards, Arnold and associates estimated that using the single arm strength test to select employees would lead to a savings of about \$5,000 per year for each employee selected. Based on employees hired, the estimated company savings were more than \$9 million a year.

## Underground Coal Mining

A job analysis showed that the work of underground coal miners was physically demanding and that the work could be represented with four work sample tests (Jackson & Osburn, 1983; Jackson et al., 1991). The first work-sample simulation test, roof bolting, measured maximum isokinetic torque and simulated straightening a steel roof bolt. The block carry test involved lifting, transporting, and placing 82-pound concrete blocks in positions commonly used to build retaining walls in the mine. The shoveling simulation test involved shoveling polyvinyl chloride from the floor over a 3.5-foot wall. Polyvinyl chloride has the same density of coal, and the task was to shovel 800 pounds at a rate consistent with the subject's fitness. The bag carry simulation test measured the number of 50-pound bags that were lifted and transported 9 feet during a 5-minute period.

The four work-sample tests and three isometric strength tests (grip, arm lift, and torso lift) (NIOSH, 1977) were administered to 25 male and 25 female subjects. The validation strategy was similar to that followed by Arnold and associates with steelworkers (Arnold et al., 1982). The correlations between the sum of the isometric strength tests and four work-sample tests ranged from 0.68 for the bag carry test to 0.91 for the roof bolting test. Multiple regression analysis showed that neither gender nor the gender-by-isometric strength interaction accounted for the additional significant variance. This showed that a common male and female regression line defined the relationship between strength and work-sample test performance.

Both exercise heart rate and rating of perceived exertion data showed that the shoveling and bag carry tests had significant aerobic components (Jackson et al., 1991). In addition to the isometric strength tests, the subject's maximal arm cranking oxygen uptake was metabolically determined. The zero-order correlations between the sum of isometric strength and the work-sample shoveling and bag carry tests were higher than the correlations found with arm  $\text{VO}_{2\text{max}}$  (ml/min). The strength correlations were 0.71 for shoveling and 0.63 for the bag carry test, compared with 0.68 and 0.46 for arm  $\text{VO}_{2\text{max}}$  (ml/min). Multiple regression analysis showed that arm  $\text{VO}_{2\text{max}}$  accounted for an additional 9 percent of shoveling variance beyond that of isometric strength but did not account for additional bag carry variance. Polynomial regression analysis showed that the relationship between

these two endurance work-sample tests and isometric strength was quadratic, not linear. Strength was more important for differentiating among work sample performance at the lowest levels.

## Chemical Plant Workers

Job analyses documented that the physically demanding tasks required of chemical and refining plants workers included cracking, opening, and closing valves (Jackson, Osburn, Laughery, & Vaubel, 1990; Osburn, 1977). Osburn (1977) developed a valve-turning work-simulation test administered on a specially developed ergometer consisting of a disc brake mechanism turned by a 12-inch valve handwheel. The unit was calibrated to a power output of 1,413.5 foot-pounds/minute. The objective of the work-sample test was to complete 250 revolutions in 15 minutes. The job analysis showed this level of work would open or close 75 percent of the emergency valves in 15 minutes.

The distribution of the valve-turning test was bimodal. Physically fit workers easily completed the 15-minute test, but the test was too demanding for many, who stopped before reaching 50 revolutions (Jackson et al., 1990). The test elicited maximal cardiovascular responses in many applicants (Osburn, 1977). This result led to a second study designed to determine whether isometric strength tests validly predicted valve-turning performance (Jackson, 1987; Jackson et al., 1992). The valve-turning work-sample test, and three isometric strength tests (grip, arm lift, and torso lift) were administered to 26 men and 25 women. The zero-order correlation between the tests was 0.82. Because of the bimodal shape of the valve-turning distribution, a logistic regression model (Pedhazur, 1997) defined the probability of completing the test by levels of isometric strength. The logistic equations and probability curves are published (Jackson et al., 1992).

In a second study, a task analysis questionnaire completed by operators at a major chemical plant identified valve cracking as the most physically demanding work task (Jackson et al., 1990). An electronic load cell measured the peak cracking torque on 217 randomly selected valves in the plant. The sampled valves included those with horizontal and vertical orientations, positioned close to the ground and overhead, those in awkward or hard to reach positions, and valves of various sizes. The results of this biomechanical job analysis showed that 100 pounds of force applied to the end of a 36-inch valve wrench generated sufficient torque to crack 93 percent of the plant valves.

A valve-cracking work-sample test simulated cracking valves in eight different ways. The eight cracking torques were obtained by varying the action (push and pull), direction (horizontal and vertical), and height (high and low). A computerized torque wrench measured the torque applied to four nuts placed in vertical and horizontal positions at two heights.

The valve-cracking test and isometric strength tests (grip, arm lift, and torso lift) were administered to 118 men and 66 women. The intercorrelations among the eight measures of valve-cracking torque were high, ranging from 0.66 to 0.89. Because of the high intercorrelations, the eight valve-cracking scores were averaged and used as the work-sample measure. The correlation between the sum of the three isometric strength tests and average valve-cracking torque was 0.65. A logistic regression equation (Pedhazur, 1997) defined a probability model for estimating the chances of generating the 100-pound criterion for levels of isometric strength. These data are published elsewhere (Jackson et al., 1992).

## Electrical Transmission Lineworkers

Doolittle and associates (Doolittle et al., 1988) developed a preemployment test for selecting electrical transmission lineworkers. The study included an extensive job analysis of electrical transmission lineworker jobs. The initial stage of the task analysis surveyed workers using scales designed to answer three questions—

1. How often was each task performed?
2. How much time was spent completing each task?
3. How physically demanding was each task for the individual?

The identified critical, physically demanding tasks were studied in detail to define the forces needed to perform them safely and efficiently. This involved defining standard anatomical movements for lifting, pushing, and hoisting; measuring the masses lifted and forces exerted; and estimating the metabolic costs of various work tasks.

Using the task analysis data, 5 strength tests that duplicated the muscular actions were selected and administered to 48 incumbents. The tests required the subject to move a weight that represented loads that linemen moved. The weights ranged from 7 to 61 kilograms. The final two tests selected were chin-ups and  $\text{VO}_2\text{max}$  estimated from bench stepping and exercise heart rate. The seven tests were combined into a single performance measure. Criterion-related validity was examined by comparing physical test performance with two criteria, supervisor ratings and accident rates. The crew chiefs confidentially evaluated each incumbent on the following six dimensions of job performance —

1. productivity,
2. working with others,
3. supervision,
4. safety,
5. physical ability, and
6. technical skills.

The correlations between the composite physical test criteria of supervisor ratings and lost work days because of on-the-job injuries averaged over 5 years were 0.59 and 0.46.

## Diver Training

Two validation studies (Gunderson, Rahe, & Arthur, 1972; Hogan, 1985) were designed to estimate successful completion of Military underwater diver training programs. Gunderson and associates (Gunderson et al., 1972) used successful completion of underwater demolition training as the criterion of performance. They found a multiple correlation of 0.54 between success defined by the completion of training and five variables, squat-jumps, pull-ups, sit-ups, body weight, and the Cornell Medical Index. Using these tests, they predicted about 70 percent of those who passed training.

Hogan (Hogan, 1985) used 46 male, naval personnel who volunteered for diver training. The first criteria was success included nine performance rating scales that reflected physical condition, swimming training, leadership potential, teamwork, and overall performance. The second criteria was successful completion of training. The predictor measures included 3 anthropometric measurements and 23 fitness tests. Hogan reported a multiple correlation of 0.63 between the average performance rating and three physical tests, 1-mile run, sit and reach, and muscular endurance measured with an arm ergometer. The multiple correlation between these three tests and successful completion of the course was 0.64. Hogan suggested that the validity coefficients were likely an overestimate because of an unfavorable ratio of the number variables and subjects (Pedhazur, 1997).

## **Demanding Military Jobs**

The U.S. Military Services examined methods of matching enlisted personnel with physically demanding jobs. The U.S. Air Force adopted a pre-induction dynamic one-repetition maximum (1-RM) strength test (Ayoub et al., 1982). The U.S. Army and U.S. Navy examined the relationship between body composition variables and physically demanding work tasks (Marriott & Grumstrup-Scott, 1992).

The U.S. Air Force developed a Strength Aptitude Test (SAT) to match the general strength abilities of individuals with the specific strength requirements of U.S. Air Force jobs filled by enlisted personnel (Ayoub et al., 1982). The U.S. Air Force SAT measures the subject's voluntary 1-RM lift to a height of 6 feet. The SAT starts with a 40-pound lift. The lift load is increased by 10 pounds until the subject reaches his or her maximum voluntary lift or a maximum weight of 200 pounds. The SAT is administered to U.S. Air Force recruits as part of their pre-induction physical examination. Each enlisted U.S. Air Force career field has a prerequisite SAT cut-score.

An area of concern expressed by the Committee on Military Nutrition Research of the Institute of Medicine, National Academy of Sciences, is the role body composition plays in physical performance. This relationship is important not only for making decisions about acceptance or rejection of recruits for the Military Service but also for retention and advancement while in the Service (Marriott & Grumstrup-Scott, 1992). Hodgdon and associates (Hodgdon, 1992) examined the relationship between body composition, fitness, and materials-handling tasks required of naval enlisted men. The two materials-handling tasks were the maximum box weight that could be lifted to elbow height and the total distance a 34-kilogram box could be carried during two, 5-minute workouts. The variables most highly correlated with maximum box lift were push-ups ( $r = 0.63$ ) and fat-free mass ( $r = 0.80$ ). The variables most highly correlated with the box carry test were push-ups ( $r = 0.56$ ), 1.5-mile run time ( $r = -0.67$ ), and fat-free mass ( $r = 0.44$ ). Fat-free mass was highly correlated with muscular strength measures, suggesting the possibility of using fat-free mass as an approximation of general strength in job assignment.

Vogel and Friedl (Vogel & Friedl, 1992) examined the relationship between body composition and absolute lifting capacity. They reported significant correlations between maximum lifting capacity and fat-free mass for male and female soldiers. Although they did not test for homogeneity of male and female regression lines, they published separate equations for men and women.

A limitation of Military testing programs is the lack of job-related materials-handling performance tests. While recognizing the need to develop content-valid tests, the Committee on Military Nutrition Research concluded that there was a direct relationship between Military materials-handling tasks and fat-free mass. In view of this relationship and the lack of job-related tests, the Military should seriously consider establishing a minimum standard for fat-free mass (Marriott & Grumpstrup-Scott, 1992). Such a recommendation might be implemented for the Military, but using body composition variables in pre-employment tests in the private sector would likely meet an immediate legal challenge.

Rayson and associates (Rayson et al., 2000a; Rayson et al., 2000b) completed a major criterion-related validation study for the British army. They examined the effectiveness of the British army's Physical Standards for Recruits (PSS(R)) in predicting criteria measuring recruit success in basic training. The PSS(R) consisted of tests measuring body mass, body composition, strength, and endurance. The criteria included—

1. four representative Military tasks (RMT) consisting of a single lift, carry, repetitive lift, and loaded march,
2. the days lost to injury and sickness during basic training,
3. degree of success of basic training, and
4. job performance ratings by self, peer, and supervisor.

The PSS(R) tests were administered to more than 1,000 recruits (770 males and 239 females) prior to starting basic training, and the army job performance criteria were obtained at the end of basic training.

The PSS(R) tests correctly predicted outcomes on the RMTs for 74.9 percent of the recruits, of which 58.7 percent were true positives and 16.2 percent were true negatives. Of the 25.1 percent misclassified, 15.5 percent were false positives and 9.6 percent were false negatives. The false negatives were those recruits predicted by the PSS(R) tests to fail the four RMTs when they did pass the tasks. Although data were not presented, the authors indicated that most of the female misclassifications were false positives, "...women being incorrectly accepted rather than incorrectly rejected from the army." A significant relationship was found between training outcome and passing the PSS(R) tests. Additionally, the PSS(R) tests were significantly related to days lost because of injury and sickness during basic training. Those recruits who failed their selection outcome lost a median of 2 days compared with no days for the recruits who passed. Although not statistically significant, the performance ratings of those who failed the selection tests were consistently lower than those who passed the tests. The authors concluded that the PSS(R) were valid, useful predictors of British army performance.

## Manual Lifting Tasks

Manual lifting tasks are common elements of many jobs. Manual lifting tasks have been studied extensively. The reason for this popularity is the large number of jobs that include materials-handling tasks and the injury risk associated with lifting. It is estimated that about 50 percent of

all industrial back injuries are caused by lifting, and about 67 percent of the injuries are caused by lifting loads that are too difficult for industrial workers (Snook et al., 1978).

An established ergonomic injury-reduction strategy is to match the worker with the demands of the lifting task. One major approach is to engineer the stress out of the task. This approach defines the lift weights that are within the physiological capacity of most industrial workers (Ayoub, 1982). The first research-based strategy used psychophysical methods to define the lift weight perceived as acceptable to 75 percent of industrial workers. Snook and associates (Snook & Ciriello, 1974; Snook & Ciriello, 1991; Snook, Irvine, & Bass, 1970) published separate standards for males and females. The maximum acceptable lift weight for females was about 50 percent of the lift weights for males. A newer strategy is the use of the NIOSH multiplicative equations (NIOSH, 1981; Waters et al., 1993) that consider several different lift difficulty parameters. The NIOSH equations extend the Snook and associates' psychophysical methodology by also using biomechanical and physiological criteria to define recommended weight of lift (RWL). The newest NIOSH equation (Waters et al., 1993) defines a RWL that would be acceptable to 75 percent of the female industrial population. Using the 75<sup>th</sup> percentile female as the RWL criterion produces a conservative estimate. The RWL for the common floor to knuckle lift at a frequency of one lift every 30 minutes, for example, is only 10 kilograms or 22 pounds (Waters et al., 1993).

The NIOSH equation focuses on job design, i.e., defining a RWL for most male (99 percent) and female (75 percent) industrial workers for all ages in the workforce. A limitation of the NIOSH equation is that it does not consider individual differences in physiological capacity of workers. Many common materials-handling tasks exceed the NIOSH equation's RWL estimates. The second ergonomic method of matching the worker with the demands of job is to select individuals with the physiological capacity to do the job with a margin of safety (Ayoub, 1982; Keyserling & al., 1980; NIOSH, 1977).

The content-validation method is often used to validate materials-handling tests. A content-valid test would be to have the applicant perform the task, e.g., lift a 90-pound jackhammer and transport it a specified distance. Although this type of test would be content valid, it has two limitations. First, it is not possible to determine one's maximum capacity. Second, motivated applicants without the physiological capacity demanded by the task place themselves at risk of injury (Ayoub, 1982). One of the first ergonomic approaches used to overcome these limitations was to use isometric strength tests that duplicated the position assumed by the worker to do the lift. These position-specific strength data were used to determine whether an applicant had sufficient strength capacity to do the work with a margin of safety (Keyserling & al., 1980; Keyserling et al., 1980).

Gilliam and Lund (2000) examined the effects on work-related injuries of physiologically matching workers to the demands of the job. Isokinetic strength was measured on 365 applicants for truck driver and dockworker jobs. The isokinetic data were used to generate a Department of Labor Dictionary of Occupational Titles strength rating. This rating was used to select applicants who matched the physical demands of the job. Of the 365 applicants, 276 matched the job demands and were hired. The 89 applicants who did not match were not hired. Those hired were significantly stronger than those who were not hired. In addition, those not hired were significantly heavier than those hired. Those not hired were 44 pounds heavier than the new hires. The injury rates of the strength-matched new hires were compared with historical data on workers matched for employment duration. The overexertion injury rates to the knees, shoulders, and back were 1.04 for the

strength-matched workers compared with 16.7 for the non-matched workers, suggesting that pre-employment screening is effective in reducing injury. Although not examined, these results also suggest that body composition may also have been a factor. A strength-weight profile of weaker and heavier versus stronger and lighter suggest a difference in percent body fat. The stronger-lighter profile is consistent with a lower percent body fat, which also might have been an injury risk factor.

Another physiological approach to matching the worker to the demands of the job is to use standard strength tests to assess an individual's physiological capacity and use regression models to define the probability of being able to complete a lift (Jackson & Sekula, 1999; Jackson, Borg, Zhang, Laughery, & Chen, 1997). This approach was used to study hospital workers involved with lifting and transporting patients. An analysis of hospital jobs documented that patient lifting was a demanding lift task (Jackson, Osburn, Laughery, Young, & Zhang, 1994). Patient lift tasks are a major source of injury to the lifter (Garg & Owen, 1992). The lift dimensions of the most common single-person patient lift were used to devise a work-sample lift test. The most common patient lift task is lifting a patient who is sitting in a chair. The simulated lift test consisted of lifting a box from a height of 53 cm to a height of 48 cm. The hand position at the start of the lift was at a height that the lifter would grab a patient sitting in a chair. The lift task consisted of lifting seven loads ranging in weight from 15 to 90 pounds. The subjects lifted those loads that were within their capacity and rated lift difficulty with Borg's CR-10 psychophysical scale (Borg, 1982; Borg, 1998). Logistic regression analysis of the data on 58 female and 33 male subjects showed that the capacity to complete a lift depended on the lifter's physiological capacity sampled by his or her isometric strength and fat-free mass. Further analyses showed that the subject's CR-10 rating of each lift was significantly correlated with isometric arm, shoulder, torso, and leg strength, and fat-free weight.

The results of the patient lift study suggested that lift weight and the physiological capacity of the lifter could be used to develop a generalized lift model. The second study examined the role of lift load, strength, and gender on psychophysical lift capacity (Jackson, 1999). A floor-to-knuckle lift test was administered to 209 men and 181 women. The task involved lifting loads ranging from 22 to 143 pounds. The subject started with a light lift load and continued to lift heavier loads until either the heaviest load was lifted or the subject failed the lift. The load increased at a linear rate of 11 pounds. After each completed lift, the subject rated the lift difficulty with Borg's CR-10 scale (Borg, 1998). The subject's physiological strength capacity was measured with basic isometric strength tests (Baumgartner & Jackson, 1999). Each subject's dynamic lift profile was defined with a power function regression equation using the completed lift weight as the independent variable and the CR-10 rating as the dependent variable. Using the power function regression equation, one lift weight and the associated CR-10 rating were randomly selected for each subject. This created a distribution of lift weights and associated psychophysical ratings ranging from very easy to the maximum within the subject's psychophysical capacity. Multiple regression provided an equation with a function to estimate psychophysical lift difficulty from lift load, strength, and the gender-by-weight load interaction. The multiple correlation for the model was 0.81, with a standard error of 1.7 CR-10 units. The derived equation provided a model that defined the psychophysical lift demands of common industrial weight loads for individuals who differed in physiological capacity.

The psychophysical modeling of industrial lift tasks not only provides evidence concerning an individual's probability of being able to complete a lift but also psychophysical stress. The psychophysical demand of a lift task is related to the risk of back injury (Herrin, 1986; Liles, 1984;



Snook et al., 1978). Lifting loads psychophysically judged to be difficult increases the risk of injury. Psychophysical ratings provide an index of relative demand for the individual. Resnik (1995) presents preliminary data showing that Borg's psychophysical rating can be interpreted by the physiological significant scale of percentage of maximum capacity. With a sample of 254 male and 354 female subjects, a correlation of 0.91 was obtained between Borg's CR-10 rating and the subject's maximum function lift capacity (Sekula, Jackson, & Laughlin, under review). Maximum functional lift capacity was the subject's percentage of maximum lift, where maximum lift represented the weight load equal to the subject's Borg psychophysical CR-10 rating of 10. A regression equation was developed to convert CR-10 ratings into the metric of percentage of max. The standard error of estimate for the linear equation was 8.5 percent max. This research could provide researchers with the capacity to interpret psychophysically defined lift loads with the well-established physiological intensity metric of percentage of maximum capacity.

---

## Summary

In *summary*, the Uniform Guidelines require validity studies to be carried out whenever there is a need to continue selection practices that lead to adverse impacts. Three types of validity studies are recognized: content-validity, criterion-related validity, and construct-validity studies. The guidelines require all validity studies to be carried out in a responsible, scientifically sound manner, and call for the use of good judgment in the implementation of selection procedures. The EEOC is waiting for developments in the field before it completely endorses construct-validity studies. A major difference in physical test validation is the use of physiological rather than psychological tests. The goal of physiological validation is to define the physiological capacity needed by a worker to perform the work demanded by the task. Principal features of the physiological validation approach are the use of a physiological metric to quantify test performance and the interpretation of validity results using relevant physiological research and theory. These data are used to develop physiologically sound cut-scores. Although numerous physical test validation studies have been completed, most are not published. The results of those published shows that physical tests can be used to select workers with the physiological capacity to do demanding jobs. Ergonomic research shows that selecting workers with the physiological capacity to do the work reduces the risk of work-related injuries.

## Endnotes

1. The probability can be estimated with the following equation:  $z = \frac{Y - \text{criterion}}{\text{standard error of estimate}}$ , where Y' is the estimated criterion score and the criterion is the desired value, in this example, 100. Once the z-score is obtained, a table of normal curves can be used to estimate the proportion of subjects that can be expected to exceed the criterion for a given strength level.
2. Personal communication between A. Jackson, University of Houston, and Dr. John Hater of the Fedex Corporation. Engineers used the power output data in Figure 5.3 to estimate expected changes in productivity produced by changes the physiological capacity of the workforce.

3. This review was initially published in 1994 by one of the authors of this chapter (Jackson, 1994) and expanded to include studies published since that time.
- 

## References

1. A.P.A. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
2. A.P.A. (1987). *Principles for the validation and use of personnel selection procedures*. Washington, DC: Division of Industrial-Organizational Psychology, American Psychological Association.
3. ACSM. (1991). *Guidelines for exercise testing and prescription*. (3<sup>rd</sup> ed.). (Vol. 4). Philadelphia, PA: Lea and Febiger.
4. Altman, D. G., Bland, J. M. (1983). Measurement in medicine: The analysis of method comparison studies. *The Statistician*, 32, 307–317.
5. Altman, D. G., Bland, J. M. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, I, 307–310.
6. Arnold, J. D., Rauschenberger, J. M., Soubel, W. G., & Guion, R. M. (1982). Validation and utility of a strength test for selecting steelworkers. *Journal of Applied Psychology*, 67, 588–604.
7. Arvey, R. D., & Faley, R. H. (1988). *Fairness in Selecting Employees*. (2<sup>nd</sup> ed.). Reading, MA: Addison-Wesley Publishing Co.
8. Åstrand, P.-O., & Rodahl, K. (1986). *Textbook of work physiology*. (3rd ed.). New York: McGraw-Hill.
9. Åstrand, P.-O., & Ryhming, I. (1954). A nomogram for calculation of aerobic capacity (physical fitness) from pulse rate during submaximal work. *Journal of Applied Physiology*, 7, 218–221.
10. Ayoub, M. A. (1982). Control of manual lifting hazards: II. Job redesign. *Journal of Occupational Medicine*, 24, 688–676.
11. Ayoub, M. M., Denardo, J. D., Smith, J. L., Bethea, N. J., Lambert, B. K., Alley, L. R., & Duran, B. S. (1982). *Establishing physical criteria for assigning personnel to Air Force jobs* (Contract No. F49620-79-C0006). Lubbock, TX: Texas Tech University.
12. Barnard, R., & Duncan, H. W. (1975). Heart rate and ECG responses of firefighters. *Journal of Occupational Medicine*, 17, 247–250.
13. Baumgartner, T. A., & Jackson, A. S. (1999). *Measurement for evaluation in physical education and exercise science*. (6<sup>th</sup> ed.). Dubuque: Wm. C. Brown.
14. Bernauer, E. M., & Bonanno, J. (1975). Development of physical profiles for specific jobs. *Journal of Occupational Medicine*, 17, 22–33.
15. Borg, G. (1982). A category scale with ratio properties for intermodal and interindividual comparisons. In H. G. Geissler & P. Petzold (Eds.), *Psychophysical judgment and the process of perception*. Berlin: VEB Deutscher Verlag der Wissenschaften.
16. Borg, G. (1998). *Borg's perceived exertion and pain scaling method*. Champaign: Human Kinetics.
17. Brooks, G., & Fahey, T. (1984). *Exercise physiology: Human bioenergetics and its applications*. New York: John Wiley and Sons.

18. Brozek, J., & Keys, A. (1951). The evaluation of leanness-fatness in man: Norms and intercorrelations. *British Journal of Nutrition*, 5, 194–206.
19. Bruce, R. A., Kusumi, F., & Hosmer, D. (1973). Maximal oxygen intake and nomographic assessment of functional aerobic impairment in cardiovascular disease. *American Heart Journal*, 85, 546–562.
20. Cady, L. D., Bishoff, D. P., O'Connell, E. R., Thomas, P. C., & Allan, J. H. (1979). Back injuries in firefighters. *Journal of Occupational Medicine*, 21, 269–272.
21. Considine, W., Misner, J. E., Boileau, R. A., Pounian, C., Cole, J., & Abbatiello, A. (1976). Developing a physical performance test battery for screening Chicago fire fighting applicants. *Public Personnel Management*, 5, 7–14.
22. Davis, J. A., & Convertino, V. A. (1975). A comparison of heart rate methods for predicting endurance training intensity. *Medicine and Science in Sports*, 7, 295–298.
23. Davis, P. O., Dotson, C. O., & SantaMaria, D. L. (1982). Relationship between simulated fire fighting and physical performance measures. *Medicine and Science in Sports and Exercise*, 14, 65–71.
24. Doolittle, T. L., Spurlin, O., Kaiyala, K., & Sovern, D. (1988). Physical demands of lineworkers. *Proceedings of the Human Factors Society, 32<sup>nd</sup> Annual Meeting*, 32, 632–636.
25. Durnin, J. V. G. A., & Passmore, R. (1967). *Energy, work and leisure*. London: Heinemann Educational Books, LTD.
26. Durnin, J. V. G. A., & Wormsley, J. (1974). Body fat assessed from total body density and its estimation from skinfold thickness: Measurements on 481 men and women aged from 16 to 72 years. *British Journal of Nutrition*, 32, 77–92.
27. EEOC. (1978). Uniform Guidelines on employment selection procedures. *Federal Register*, 43 (38289–28309).
28. Fleishman, E. A. (1964). *The structure and measurement of physical fitness*. Englewood Cliffs, NJ: Prentice-Hall.
29. Foster, C., Jackson, A. S., & Pollock, M. L. (1984). Generalized equations for predicting functional capacity from treadmill performance. *American Heart Journal*, 107, 1229–1234.
30. Garg, A., & Owen, B. (1992). Reducing back stress to nursing personnel: An ergonomic intervention in a nursing home. *Ergonomics*, 35, (11), 1353–1375.
31. Gettman, L. R. (1993). Chapter 19 Fitness Testing. In *Resource manual for guidelines for exercise testing and prescription* (2<sup>nd</sup> ed., pp. 229–246). Philadelphia: Lea & Febiger.
32. Gilliam, T. B., & Lund, S. J. (2000). Injury reduction in truck driver/dock workers through physical capability new hire screening. *Medicine and Science in Sports and Exercise*, 32, (5) (Supplement), S126.
33. Gunderson, E. K. E., Rahe, R. H., & Arthur, R. J. (1972). Prediction of performance in stressful underwater demolition training. *Journal of Applied Psychology*, 56, pp. 430–432.
34. Herrin, G. D., Jaraiedi, M., Anderson, C. K. (1986). Prediction of overexertion injuries using biomechanical and psychophysical models. *American Industrial Hygiene Association Journal*, 47, 322–330.
35. Hodgdon, J. A. (1992). Body composition in the military services: standards and methods. In B. M. Marriott & J. Grumstrup-Scott (Eds.), *Body composition and physical performance: Applications for the military services* (pp. 57–70). Washington, DC: National Academy Press.
36. Hogan, J. (1985). Tests for success in diver training. *Journal of Applied Psychology*, 70, 219–224.
37. Hogan, J. (1991a). The structure of physical performance in occupational tasks. *Journal of Applied Psychology*, 76, 495–507.

38. Hogan, J. C. (1991b). Chapter 11 Physical Abilities. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of Industrial and Organizational Psychology* (2<sup>nd</sup> ed., Vol. 2, pp. 743–831). Palo Alto, CA: Consulting Psychologist Press, Inc.
39. Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression*. New York: John Wiley & Sons.
40. Hulley, S. B., Cummings, S. R. (1988). *Designing clinical research*. Baltimore, MD: Williams & Wilkins.
41. Hunter, J. E., Schmidt, F. L., & Hunter, R. (1979). Differential validity of employment tests by race: A comprehensive review and analysis. *Psychological Bulletin*, 86, 721–735.
42. Jackson, A. S. (1987). *Validity of isometric strength tests for predicting work performance of refinery workers*. Houston: Shell Oil Company.
43. Jackson, A. S. (1989). Chapter 9: Application of regression analysis to exercise science. In M.J. Safrit & T.M. Wood (Eds.), *Measurement concepts in physical education and exercise science*. Champaign: Human Kinetics.
44. Jackson, A. S., Blair, S. N., Mahar, M. T., Wier, L. T., Ross, R. M., Stuteville, J. E. (1990). Prediction of functional aerobic capacity without exercise testing. *Medicine and Science in Sports and Exercise*, 22, 863–870.
45. Jackson, A. S. (1994). Chapter 3 Preemployment Physical Evaluation. *Exercise and Sport Science Reviews*, 22, 53–90.
46. Jackson, A. S., Osburn, H. G., Laughery, K. R., Sekula, B. K. (1998). *Revalidation of methods for pre-employment assessment of physical abilities at Shell Western Exploration and Production, Inc., and CalResources LLC*. Houston, TX: University of Houston.
47. Jackson, A. S., Sekula, B. (1999). The influence of strength and gender on defining psychophysical lift capacity. *Proceedings of the Human Factors and Ergonomics Society*, 43, 723–727.
48. Jackson, A. S., Borg, G., Zhang, J. J., Laughery, K. R., & Chen, J. (1997). Role of physical work capacity and load weight on psychophysical lift ratings. *International Journal of Industrial Ergonomics*, 20, 181–190.
49. Jackson, A. S., & Osburn, H. (1983). *Preemployment physical test development for coal mining technicians*. Technical Report to Shell Oil Co. Houston: Shell Oil Company.
50. Jackson, A. S., Osburn, H. G., & Laughery, K. R. (1984). Validity of isometric strength tests for predicting performance in physically demanding jobs. *Proceedings of the Human Factors Society 28<sup>th</sup> Annual Meeting*, 28, 452–454.
51. Jackson, A. S., Osburn, H. G., Laughery, K. R., & Vaubel, K. P. (1990). *Validation of physical strength tests for the Texas City Plant — Union Carbide Corporation*. Houston: Center for Psychological Services.
52. Jackson, A. S., Osburn, H. G., Laughery, K. R., & Vaubel, K. P. (1992). Validity of isometric strength tests for predicting the capacity to crack, open and close industrial valves. *Proceedings of the Human Factors Society 36<sup>th</sup> Annual Meeting*, 1, 688–691.
53. Jackson, A. S., Osburn, H. G., Laughery, K. R., & Young, S. L. (1993a). *Validation of physical strength tests for the Federal Express Corporation*. Houston: Center of Applied Psychological Services, Rice University.
54. Jackson, A. S., Osburn, H. G., Laughery, K. R., Young, S. L., & Zhang, J. J. (1994). *Patient lifting tasks at Methodist Hospital*. Houston: Center of Applied Psychological Services, Rice University.

55. Jackson, A. S., Osburn, H. G., & Laughery, S., K.R. (1991). Validity of isometric strength tests for predicting endurance work tasks of coal miners. *Proceedings of the Human Factors Society 35<sup>th</sup> Annual Meeting*, 1, 763–767.
56. Jackson, A. S., & Pollock, M. L. (1978). Generalized equations for predicting body density of men. *British Journal of Nutrition*, 40, 497–504.
57. Jackson, A. S., Pollock, M. L., & Ward, A. (1980). Generalized equations for predicting body density of women. *Medicine and Science in Sports and Exercise*, 12, 175–182.
58. Jackson, A. S., Zhang, J. J., Laughery, K. R., Osburn, H. G., & Young, S. L. (1993b). *Final report: Patient lifting tasks at Methodist Hospital*. Houston: Center of Applied Psychological Services, Rice University.
59. Jeanneret & Associates, I. (1999). *Evaluation of physical ability and written tests for entry-level firefighters*. St. Paul, MN: City of St. Paul Fire Department.
60. Kannel, W. B., McGee, D., & Gordon, T. (1976). A general cardiovascular risk profile: The Framingham study. *American Journal of Cardiology*, 38, 46–51.
61. Keyserling, W. M., et al. (1980). Establishing an industrial strength testing program. *American Industrial Hygiene Association Journal*, 41, 730–736.
62. Keyserling, W. M., Herrin, G. D., & Chaffin, D. B. (1980). Isometric strength testing as a means of controlling medical incidents on strenuous jobs. *Journal of Occupational Medicine*, 22, 332–336.
63. Landy, F. J., & Investigator, P. (1992). *Alternatives to chronological age in determining standards of suitability for public safety jobs: Volume I: Technical report*. The Pennsylvania State University: Center for Applied Behavioral Sciences.
64. Lemon, P., & Hermiston, R. T. (1977). The human energy cost of firefighting. *Journal of Occupational Medicine*, 19, 558–562.
65. Liles, D. H., Deivanayagam, S., Ayoub, M. M., Mahajan, P. (1984). A job severity index for the evaluation and control of lifting injury. *Human Factors*, 26, 683–693.
66. Manning, J., & Griggs, T. (1983). Heart rates in fire fighters using light and heavy breathing equipment: Similar near-maximal exertion in response to multiple work load conditions. *Journal of Occupational Medicine*, 25, 215–218.
67. Marriott, B. M., Grumstrup-Scott, J. (Eds.) (1992). *Body composition and physical performance: Application for the military services*. Washington, DC: National Academy Press.
68. McArdle, W. D., Katch, F. I., & Katch, V. L. (1991). *Exercise physiology: Energy, nutrition, and human performance*. (3<sup>rd</sup> ed.). Philadelphia: Lea & Febiger.
69. McArdle, W. D., Katch, F. I., & Katch, V. L. (1996). *Exercise physiology: Energy, nutrition, and human performance*. (4<sup>th</sup> ed.). Philadelphia: Lea & Febiger.
70. Meyers, D. C., Gebhardt, D. L., Crump, C. E., & Fleishman, E. A. (1984). *Factor analysis of strength, cardiovascular endurance, flexibility, and body composition measures* (Tech. Rep. R83–9). Bethesda, MD: Advanced Research Resources Organization.
71. NIOSH. (1977). *Preemployment strength testing*. Washington, DC: U.S. Department of Health and Human Services.
72. NIOSH. (1981). *Work practices guide for manual lifting*. Washington, DC: U.S. Department of Health and Human Services.
73. O'Connell, E., Thomas, P., Caddy, L., & Karwasky, R. (1986). Energy costs of simulated stair climbing as a job-related task in fire fighting. *Journal of Occupational Medicine*, 28, 282–284.

74. Osburn, H. G. (1977). *An investigation of applicant physical qualifications in relation to operator tasks at the Deer Park Manufacturing Complex*. Houston, TX: Employee Relations, Shell Oil Company.
75. Passmore, R., & Durnin, J. V. G. A. (1955). Human energy expenditure. *Physiological Review*, 35, 801–840.
76. Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction*. (3<sup>rd</sup> ed.). New York: Harcourt Brace College Publishers.
77. Pollock, M. L., Bohannon, R. L., Cooper, K. H., Ayres, J. J., Ward, A., White, S. R., & Linnerud, A. C. (1976). A comparative analysis of four protocols for maximal treadmill stress testing. *American Heart Journal*, 92, 39–42.
78. Rayson, M., Pynn, H., Rothwell, A., Nevill, A. (2000a). Validation of job-related fitness tests. *Medicine and Science in Sports and Exercise*, 32, (5) (Supplement), S126.
79. Rayson, M. P., Pynn, H., Rothwell, A., Nevill, A. (Ed.). (2000b). *The development of physical selection procedures for the British Army. Phase 3 Validation*. London: Taylor and Francis.
80. Reilly, R. R., Zedeck, S., & Tenopir, M. L. (1979). Validity and fairness of physical ability tests for predicting craft jobs. *Journal of Applied Psychology*, 64, 267–274.
81. Resnik, M. L. (1995). The generalizability of psychophysical ratings in predicting the perception of lift difficulty. *Proceedings of the Human Factors Society*, 679–682.
82. Rummel, R. J. (1970). *Applied factor analysis*. Evanston: Northwestern University Press.
83. Sekula, B., Jackson, A. S., Laughlin, M. S. (Under Review). *Measuring percentage of maximal functional lift capacity: Calibration of Borg's CR-10 scale*.
84. Snook, S. H., Campanelli, R. A., & Hart, J. W. (1978). A study of three preventive approaches to low back injury. *Journal of Occupational Medicine*, 20, 478–481.
85. Snook, S. H., & Ciriello, B. M. (1974). Maximum weights and work loads acceptable to female workers. *Journal of Occupational Medicine*, 16, 527–534.
86. Snook, S. H., & Ciriello, V. M. (1991). The design of manual handling tasks: revised tables of maximum acceptable weights and forces. *Ergonomics*, 34, 1197–1213.
87. Snook, S. H., Irvine, C. H., & Bass, S. F. (1970). Maximum weights and work loads acceptable to male industrial workers. *American Industrial Hygiene Association Journal*, 31, 579–586.
88. Sothmann, M. S., Saupe, K., Jasenor, D., & Blaney, J. (1992). Heart rate response of firefighters to actual emergencies. *Journal of Occupational Medicine*, 34, 797–800.
89. Sothmann, M. S., Saupe, K. W., Jasenor, D., Blaney, J., Donahue-Fuhrman, S., & Woulfe, T. (1990). Advancing age and the cardiorespiratory stress of fire suppression: Determining a minimum standard for aerobic fitness. *Human Performance*, 3, 217–236.
90. Vogel, J. A., & Friedl, K. E. (1992). Army Data: Body composition and physical capacity. In B. M. Marriott & J. Grumstrup-Scott (Eds.), *Body composition and physical performance: Applications for the military services* (pp. 89–104). Washington, DC: National Academy Press.
91. Waters, T. R., Baron, S. L., Piacitelli, L. A., Anderson, V. P., Skov, T., Haring-Sweeney, M., Wall, D. K., Fine, L. J. (1999). Evaluation of the revised NIOSH lifting equation. *Spine*, 24, 386–395.
92. Waters, T. R., Putz-Anderson, V., Garg, A., & Fine, L. J. (1993). Revised NIOSH equation for the design and evaluation of manual lifting tasks. *Ergonomics*, 7, 749–766.
93. Wellens, R. I., Roche, A. F., Khamis, H. J., Jackson, A. S., Pollock, M. L., & Siervogel, R. M. (1996). Relationships between body mass index and body composition. *Obesity Research*, 4, 35–44.

94. Wilmore, J. H., & Costill, D. L. (1994). *Physiology of sport and exercise*. Champaign, IL: Human Kinetics.
95. Wilmore, J. H., & Davis, J. A. (1979). Validation of a physical abilities field test for the selection of state traffic officers. *Journal of Occupational Medicine*, 21, 33-40.





# Chapter 6

## Establishing Performance Standards

Deborah L. Gebhardt, Ph.D.  
Human Performance Systems, Inc.  
Beltsville, MD

---

### Abstract

In the employment setting, test scores may be used to determine and predict acceptable job performance. This chapter focuses on the methodologies used to establish passing scores for tests that identify individuals who are able to perform, or be trained to perform, the essential job tasks. The methodologies discussed focus on the data generated when content and criterion-related validity strategies are used to identify legally defensible passing scores.

The design of effective criterion measures that assess and differentiate levels of job performance are discussed. These criterion data, along with test scores and validity coefficients, are used to formulate passing scores that identify successful and unsuccessful candidates. Methods are explained that assess whether a passing score maximizes correct testing decisions, while also minimizing testing errors. These methods include expectancy tables, contingency tables, and Taylor-Russell tables. The use of ergonomic and normative data for setting standards is also discussed.

Issues related to test fairness and adverse impact, and their integration with legal requirements, are outlined. The effect of basic physiological tests (e.g., aerobic capacity, strength tests) and job simulations on the reduction of adverse impact is also shown by using comparisons from a variety of physically demanding jobs. Finally, the computation of test fairness is described, along with its relationship to adverse impact and test utility.

---

## Introduction

Assessment procedures may be used for a variety of purposes such as selection, classification, and placement. When assessment procedures are used in a selection setting, an applicant is typically accepted or rejected based on test scores. Single or multiple test scores can be used to determine whether an applicant has the ability to perform the job tasks. For example, an individual applying for a police officer position may take a physical performance test that requires attainment of a specific passing score. All individuals who achieve that score are considered acceptable and eligible for hire, while those who do not are eliminated from the selection process.

In contrast, test scores used for classification and placement allow for assignment of individuals to different groupings or categories (e.g., levels of aerobic capacity). A woman's aerobic capacity, for example, can be used to categorize her fitness level using normative data. A 39-year-old woman who achieves a  $\text{VO}_2\text{max}$  score of  $35 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$  would be classified in the 60<sup>th</sup> percentile, or "above average," while a 20-year-old woman with the same score would be classified between the 40<sup>th</sup> and 50<sup>th</sup> percentiles or "below average" fitness level (American College of Sports Medicine, 2000; Golding, Myers, & Sinning, 1989). These types of classifications can be used to assign individuals to work or training groups, but can not be used for selection purposes (Civil Rights Act of 1991).

This chapter focuses on how to identify specific test scores for use in making employment decisions. This process includes identifying minimum job requirements for new hires, for promotion to jobs with unique physical demands, and for retention of incumbent personnel. The discussion focuses on establishing specific test scores that are predictive of acceptable job performance. The specific minimum scores that identify individuals who pass a test are called by a variety of names: cutoff score, cut-score, passing score, or performance standard. The term passing score is used in this chapter to refer to a test score that is indicative of acceptable performance.

---

## Requirements for Identifying Passing Scores

The purpose of using passing scores or standards is to identify individuals who are able to perform, or be trained to perform, essential job tasks. Failure to meet the established score can result in an individual not being hired or promoted, or having to retake the test. The Equal Employment Opportunity Commission (EEOC) Uniform Guidelines on Employee Selection Procedures (1978) indicate that passing scores "...should normally be set so as to be reasonable and consistent with normal expectation of acceptable proficiency within the work force" (Section 5H). Although the EEOC Uniform Guidelines (1978) do not specifically define the terms "reasonable and consistent" (Section 5H), they do allow organizations to establish and interpret passing scores in relation to their specific goals such as staff utilization, growth, and profit.

Other Federal statutes, such as the Civil Rights Act of 1964 (Title VII) and 1991 (CRA 1964, 1991) and the Americans with Disabilities Act of 1990, prohibit discrimination against protected groups (e.g., minorities). The CRA of 1964 prohibits discrimination based on race, color, gender, national origin, or religion. The result of Title VII of the CRA of 1964 is that tests must be valid

and fair for legally protected groups. The CRA of 1991 further stipulates that separate passing scores for subgroups (e.g., race, gender) shall not be used for assessments that affect employment standing (e.g., selection, promotion). These statutes do not state that the tests cannot have adverse impact. Rather, tests that have adverse impact are acceptable if they are valid and job relevant.

Prior to determining a passing score for a test, one must ensure the test is valid. That is, does the test battery measure “what it is purported to measure” (Anastasi, 1996)? Three methods or experimental designs can be used to establish validity: content, criterion-related, or construct validity (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999).

Content validity requires that the test or test components sample the content of the job. The test contains simulations of essential job tasks identified in the job analysis. For example, a dock-worker test may involve stacking and unstacking cargo in a specified time frame.

A criterion-related validity strategy empirically assesses the extent to which a test is related to a measure of job performance called a criterion measure. Two types of criterion-related validity designs can be used. The first, concurrent validity, correlates existing workers’ (incumbents) scores on the test (predictor) with a measure of job performance (criterion). The second design, predictive validity, involves administering the test to job applicants and obtaining a measure of job performance at a future date. Due to practicality issues, the concurrent validity design is used most often.

Finally, construct validity measures the extent to which the test measures the trait or ability (e.g., aerobic capacity) identified in the job analysis as critical to effective job performance. For example, muscular strength may be considered a construct and can be measured with a variety of strength tests. For construct validity to be present, the correlation between the test used to assess muscular strength and other tests known to assess the same ability should be high. Further, correlations between muscular strength tests and tests of other abilities (e.g., equilibrium) should be lower.

---

## Methods for Determining Passing Scores or Standards

A variety of methods are used to identify valid passing scores. The methodology employed depends on the type of validity strategy used (e.g., content, criterion-related). The methods are based on decision theory and establishing the usefulness of the testing procedure. Although there are a variety of techniques, only the following will be addressed in this chapter.

1. Expectancy tables
2. Contingency tables
3. Taylor Russell tables
4. Normative data
5. Ergonomic data

To use Taylor Russell, expectancy, or contingency tables, two or three of the following parameters must be available—

1. test scores,
2. measures of job performance or criterion measure data, and
3. a validity coefficient, which is the correlation coefficient between the test score(s) and the criterion measure.

To illustrate the use of these three types of tables, a criterion-related validity approach is used to obtain the needed information. Use of ergonomic and normative data is addressed separately.

## Formulation of Criterion Measures

Prior to determining the usefulness of a test and identifying a passing score when using Taylor-Russell, expectancy, or contingency tables, a criterion measure or job performance measure must be developed. The criterion measure is as important as the test in a criterion-related validity study. Criterion measures should represent the important components of the job (e.g., shovel gravel for 30 minutes) and be measurable. They must be reliable and discriminate among different levels of performance across individuals. Finally, the criterion measure should define a level of acceptable job performance.

There are a variety of reliable criterion measure formats. Two commonly used criterion measures are ratings of job performance and work samples that simulate actual job tasks (Gebhardt, Baker, & Sheppard, 1998b; Gebhardt, Baker, Sheppard, & de Miranda, 1994; Landy & Farr, 1980, 1983).

**Ratings of Job Performance**—Two frequently used rating formats are graphic scales and behaviorally anchored rating scales. These scales are constructed to discriminate among different levels of performance. Graphic rating scales consist of a set of scale points (e.g., 1–10) with corresponding generic, qualitative descriptors on which the rater marks the performance level for each of several job behaviors (e.g., ability to lift 50-pounds repetitively). Figure 6.1 shows examples of two graphic rating scales.

<b>Scale 1</b>	
5	= Above average
4	= Slightly above average
3	= Average
2	= Below average
1	= Poor
<b>Scale 2</b>	
5	= Greatly exceeds job requirements
4	= Exceeds job requirements
3	= Meets job requirements
2	= Meets minimal job requirements, with assistance
1	= Does not meet job requirements

Journal of Applied Psychology, 1939, Vol. 23, pp. 565–578.

**Figure 6.1** Graphic rating scale examples

Behaviorally anchored rating scales or BARS are typically developed using critical incidents in the job that identify behaviors that are indicative of good, average, and poor performance (Flanagan, 1954; Smith & Kendall, 1963). Figure 6.2 provides an example of a BARS that lists a task and levels of performance for paramedics carrying a patient in a stair chair (Gebhardt, Baker, Sheppard, & Leonard, 1999). Levels 1, 2, and 3 were identified by supervisors as unacceptable levels of performance for this task. Therefore, raters (supervisors and/or peers) who assign a rating of “3” considered the person unacceptable.

Task Carry patient down stairs in stair chair.	
6	= Able to descend 3 floors of stairs with 250-lb. patient in stair chair.
5	= Able to descend 3 floors of stairs with 220-lb. patient in stair chair.
4	= Able to descend 3 floors of stairs with 200-lb. patient in stair chair.
3	= Able to descend 3 floors of stairs with 180-lb. patient in stair chair.
2	= Able to descend 3 floors of stairs with 160-lb. patient in stair chair.
1	= Able to descend 3 floors of stairs with 150-lb. patient in stair chair.

**Figure 6.2 BARS scale example**

A derivative of the BARS, called the behavioral observational scales (BOS), measures the frequency of desired behaviors (e.g., able to lift patient loaded gurney into ambulance) (Latham & Wexley, 1977). Similar to BARS, critical incidents and job behaviors are obtained and categorized. Supervisors and/or peers use a frequency-based scale to rate each behavior. Raters read the behavior and determine what percentile of the time the individual applies the behavior successfully. Figure 6.3 presents an example of a BOS and a list of behaviors that were rated. Research has shown that use of more complicated rating scales such as BARS is occasionally better than uncomplicated scales such as BOS or graphic (Giffin, 1989). However, both BARS and graphic scales have been used with equal success in the physical domain to validate basic ability and job simulation tests (Gebhardt et al., 1998a & b, 1999).

BOS Scale	
5	= Engages in behavior 95–100% of the time
4	= Engages in behavior 85–94% of the time
3	= Engages in behavior 75–84% of the time
2	= Engages in behavior 65–74% of the time
1	= Engages in behavior less than 65% of the time
Example of Behaviors Rated with BOS	
1.	<u>4</u> Able to lift patient from stretcher to ambulance.
2.	<u>4</u> Carries equipment to accident scene (quick response bag, oxygen tank, drug box)
3.	<u>2</u> Able to descend 2 floors of stairs while carrying patient in a stair chair.

**Figure 6.3 BOS example**

Ratings of performance by multiple supervisors, peers, or other job-knowledgeable individuals using these scales have been shown to be reliable measures ( $r = 0.50 - 0.75$ ) of job performance (Gebhardt et al., 1994, 1998b; Mumford, 1983). Although supervisor ratings of workers' behaviors are frequently used in validation studies, peer ratings have been shown to be equally effective predictors of job success (Cederbloom, 1989; Gebhardt et al., 1998a, Gebhardt, Schemmer, & Crump, 1985). These ratings are effective because both types of raters have observed job performance and are familiar with what constitutes effective job performance of critical job behaviors.

When using ratings of performance, the researcher must be cognizant of the effects of rating errors. There are three main types of errors: halo, leniency, and central tendency (King, Hunter, & Schmidt, 1980).

The *halo* effect indicates the rater has an overall favorable or unfavorable impression of the worker and rates all job tasks or behaviors in a manner consistent with this impression. *Leniency* is generally defined as assigning all favorable ratings due to an unwillingness to assign different ratings for each task or behavior. Similarly, the error of *central tendency* occurs when the rater uses only the middle of the scale and does not assign ratings that are at the scale extremes (e.g., poor, excellent). To use job performance ratings successfully, these errors must be eliminated or reduced by defining specific job traits or behaviors, forcing discrimination in ranking workers, and training all raters (Anastasi, 1996).

**Work Sample Simulations**—Work sample criterion measures are composed of essential job tasks that simulate the work environment (e.g., equipment). These simulations must be constructed to allow for quantitative analysis of the simulated tasks. Further, the reliability of the simulation should be determined. An example of a measurable work simulation in the tire manufacturing industry consisted of moving tires from a simulated inspection station to a storage facility in a specific time frame (Crump, Gebhardt, Guerette, & Wertheimer, 1985). The type and number of tires inspected and moved in the simulation were based on the factory production schedule. The acceptable level of performance on the simulation was determined from production data that indicated the number of tires moved in a specific time frame. Test-retest reliabilities for this criterion measure and similar ones in other studies ranged from .79 to .98 (Crump et al., 1985; Gebhardt et al., 1999).

Regardless of the type of criterion measures employed (e.g., ratings, work samples), it is imperative that they be reliable and measure a representative sample of the job behaviors. Further, an acceptable level of performance on the criterion measure must be determined in order to use contingency and Taylor-Russell tables.

## Identification of Acceptable Job Performance

To establish a passing score, a point on the criterion measure that differentiates between acceptable and unacceptable job performance, must be identified. One method of determining acceptable job performance is to define an acceptable level of job performance in the actual criterion measure. Using this method, a supervisor rating scale may define levels of performance as—

1. below job requirements,
2. meets job requirements, or
3. exceeds job requirements.

If a supervisor uses a three-point scale to evaluate 10 work behaviors, where acceptable performance is defined as level “2,” a worker would need a total summed score of 20 to be considered as having met job requirements (i.e., 10 tasks x 2 [meets job requirements] = 20). Note that the supervisor could rate the worker as a “2” on some tasks and a “1” or “3” on others, and the worker would still have a total of 20, indicating that the worker’s performance was acceptable.

Supervisor and peer ratings can also be used to identify acceptable and unacceptable performance of a work sample criterion. For example, a work sample criterion-measure was developed that simulated a series of critical firefighter tasks (Sothmann, Gebhardt, Baker, Costello, & Sheppard, 1995). To determine the acceptable and unacceptable performance levels on the simulation, six videotapes were generated to show six different paces of movement. The paces were established based on statistical criteria (e.g., mean, one standard deviation below mean) from a sample of incumbent firefighters. To determine the minimally acceptable pace for completing the simulation, a rating instrument was designed to allow the raters to view each pace and determine whether the pace was acceptable or unacceptable at an actual fire. A random counterbalanced design was used for presentation of the paces to the raters. The results indicated that raters were able to correctly rank the paces from fast to slow, and they demonstrated a high level of agreement on the paces that were acceptable and unacceptable. Further, the minimally acceptable pace was identified and used as the measure of acceptable job performance. This pace was then used to justify the passing score for a firefighter selection test.

Another method of defining acceptable job performance is through the use of ergonomic and production data. Gebhardt, Schemmer, and Crump (1985) completed a study in the longshore industry in which one of the essential tasks involved securing 40-foot containers to the deck of a ship with long metal rods and turnbuckles. This task required affixing the rods to the container and securing them by attaching and turning a turnbuckle until it locked the rods in place. This task was simulated as a criterion measure in a validation study. The level of acceptable job performance for the tightening and loosening of the turnbuckle was obtained by measuring the force required to “break” a tightened turnbuckle and the number of turns on a turnbuckle needed to secure a rod to a container. Individuals able to exert the required force and turn the turnbuckle the required number of times in the specified time period exhibited acceptable job performance. The time period and number of turnbuckles attached were established from ergonomic data. These data were obtained for a variety of sizes of container ships and included the time required to lash (attach rods and turnbuckles) all containers to the ship deck, the number of longshore workers assigned to a lashing operation, the number of containers placed in a stack, and the number of container stacks lashed per row.

## **Decision Models for Investigating Test Effectiveness and Setting Passing Scores**

The purpose of using tests and other assessment procedures is to predict future job performance by identifying individuals who can perform essential job tasks in a safe and effective manner. Individuals who pass a test should have a high probability of job success in an organization. If the passing score is set too high, the test will have adverse impact on protected groups and reject indi-

viduals who are qualified for the job. If the passing score is set too low, very few candidates will be screened out, allowing unqualified individuals to be hired and diminishing the test effectiveness.

Passing scores need to meet multiple goals. First, a passing score should be at a level that reflects acceptable job performance. Second, a passing score should maximize correct testing decisions and minimize testing errors. Third, the adverse impact of the test and its passing score must be considered. For example, a passing score set at a low level may have little or no adverse impact. However, a passing score set at a high level may result in adverse impact. Fourth, the passing score should be set at a level that provides test utility. The primary purpose of a selection test is to identify applicants who will be successful on the job and screen out those who will not. What makes the task of setting a passing score more difficult is that these goals are in conflict with one another. Thus, the challenge of setting an accurate passing score is the integration of these goals with the needs of the hiring organization.

For tests with demonstrated validity and reliability, several methods can be used to investigate the test's usefulness and, in turn, set a passing score at a point that identifies future successful employees. Three of these methodologies are described below: (1) Taylor-Russell tables, (2) expectancy tables, and (3) contingency or sensitivity tables.

**Taylor-Russell Tables** — The Taylor-Russell tables allow a researcher to estimate the percentage of new employees who will be successful on the job if the test is used (Taylor & Russell, 1939). To use these tables one must know the following—

1. validity coefficient for the test (e.g.,  $R=.70$ ),
2. selection ratio, and
3. base rate.

The *validity coefficient* (correlation between test and criterion measure) can be obtained through a criterion-related validity study or by estimating the test validity through past research for similar methods and studies (e.g., validity generalization). The *selection ratio* is the percentage of applicants an organization must hire to fill vacant jobs in relation to the total number of applicants (i.e., number of openings/number of applicants for the position).

Calculation of the *base rate* involves two steps. First, the number of current employees who were hired prior to the use of the test is determined. Second, the percentage of these current employees who demonstrate successful job performance is computed. One method to obtain this number is to set a criterion for successful job performance and determine which employees are above it and below it. For example, if a tire manufacturer requires that a tire builder produce 50 tires per shift, those employees who produce 50 or more are considered successful and those who produce fewer are unsuccessful. Therefore, if there are 100 tire builders and 60 produce 50 or more tires per day, the base rate is .60 (i.e.,  $60/100 = .60$ ).

After the validity, selection ratio, and base rate are determined, the Taylor-Russell tables are examined. For illustrative purposes, the Taylor-Russell .60 base rate table is presented in Table 6.1. This .60 table is used because 60% of existing workers who did not take the test are performing successfully on the job. Using the tire builder example, we assume the physical performance test had a test validity coefficient of .70 and the selection ratio was .50. Based on these data, Table 6.1 shows that 84% (.84) of the applicants who pass the test and are hired would be successful on the job. This



**Table 6.1 Example of Taylor-Russell table for base rate .60**

Validity	Selection Ratio										
	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95
.00	.60	.60	.60	.60	.60	.60	.60	.60	.60	.60	.60
.10	.68	.67	.65	.64	.64	.63	.63	.62	.61	.61	.60
.20	.75	.73	.71	.69	.67	.66	.65	.64	.63	.62	.61
.30	.82	.79	.76	.73	.71	.69	.68	.66	.64	.62	.61
.40	.88	.85	.81	.78	.75	.73	.70	.68	.66	.63	.62
.50	.93	.90	.86	.82	.79	.76	.73	.70	.67	.64	.62
.60	.96	.94	.90	.87	.83	.80	.76	.73	.69	.65	.63
.70	.99	.97	.94	.91	.87	<b>.84</b>	.80	.75	.71	.66	.63
.80	1.00	.99	.98	.95	.92	.88	.83	.78	.72	.66	.63
.90	1.00	1.00	1.00	.99	.97	.94	.88	.82	.74	.67	.63
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.86	.75	.67	.63

**Expectancy Tables**—An expectancy table shows the level of job performance (i.e., criterion measure) expected based on a particular physical performance test score. This table indicates the probability of different job performance outcomes for individuals obtaining specific test scores. Expectancy table values are determined from data generated in a criterion-related validity study in which workers are evaluated on the test and job performance. This approach allows one to predict a future worker's level of job performance when various minimum passing scores are applied to an applicant population.

Expectancy tables are generated by ranking the workers on a single test score or a combination (e.g., sum) of multiple test scores. The distribution of test scores is divided into 10 equal intervals, with each interval containing approximately the same number of workers. The mean criterion score (i.e., job performance) at each 10% interval distribution of test scores is calculated to produce an expectancy table.

A sample expectancy table is presented in Table 6.2. This table lists (1) the mean test score at each 10% interval, (2) the percent of incumbents who attained the mean test score or higher at each interval, and (3) the expected job performance (criterion measure) for the incumbents who met or

**Table 6.2 Combined physical performance test battery<sup>1</sup> expectancy table**

Combined Physical Performance Test Battery Expectancy Table			
% Level	Test Score	% Incumbents Score Higher than Test Mean <sup>1</sup> (n = 87)	Job Performance
	0 0	1000	32 0
10	431.7	90.8	38.1
20	455.1	80.5	45.3
30	483.2	70.1	53.1
40	518.7	59.2	58.2
50	550.0	50.6	63.2
60	569.5	41.4	67.3
70	595.7	31.0	71.6
80	605.9	21.8	75.8
90	647.3	10.3	78.9

1. Unequal percent of incumbents per interval due to unequal n.

exceeded the test score. The mean test score for the first 10% interval is 431.7. At this level 90.8% of the incumbent workers achieved a test score of 431.7 or greater. This means that 9.2% (i.e.,  $100\% - 90.8\% = 9.2\%$ ) of the incumbents failed the test. Similarly, if a minimum passing score of 455.1 was selected, 80.5% of the incumbents would pass the test and 19.5% would fail.

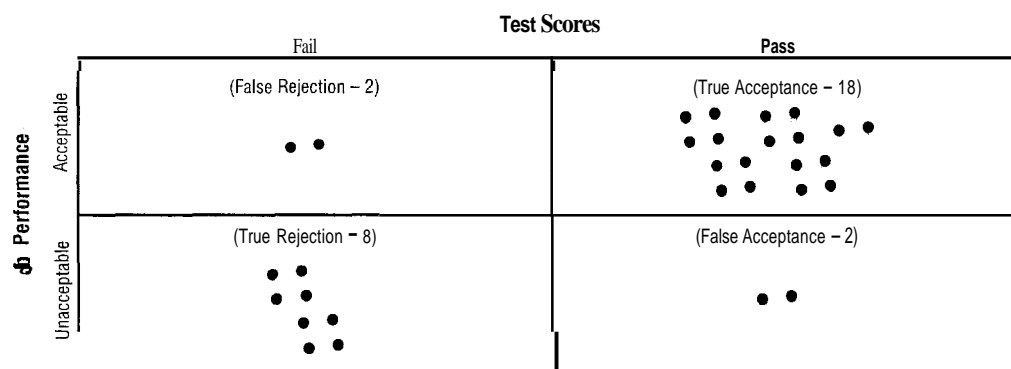
Table 6.2 also shows the impact of test performance on job performance. For example, the job performance measure used in this table is supervisor/peer ratings of performance for 10 essential tasks in which a total of 30 points defines meeting job requirements. If an applicant attained a test score of 431.7, the expected job performance would be 38.1 or 8.1 points above the minimum acceptable level. If a score of 455.1 were achieved, the average job performance would be 45.3 points or 15.3 points above the minimum.

Of interest is the increase in job performance at different levels. In Table 6.2, the mean job performance for the entire sample of workers was 32 points, which is an acceptable job performance score. However, there were workers in this sample who were unacceptable. Therefore, the increase in job performance from the mean (32) to a test score of 431.7 (10% level) was 6.1 points ( $38.1 - 32.0$ ). Similarly, the increase in job performance between a test score of 431.7 (10% level) and 455.1 (20% level) was 7.2 points. These data indicate that the 20% level score provides a greater increase in predicted job performance, and therefore, a more successful worker than the 10% level. However, it is necessary to investigate whether the 20% level results in more or less accurate decisions related to individuals who pass the test and have acceptable job performance and those who fail the test and are unacceptable on the job. Determining the accuracy of a potential passing score is accomplished through the use of contingency tables.

**Contingency Tables** — Contingency tables are used to determine the accuracy of a proposed passing score in classifying workers as acceptable and not acceptable based on their test scores. If a testing procedure were perfect, all individuals who passed the test would be successful on the job and

Figure 6.4 illustrates a distribution of test scores in relation to job success. The four quadrants contain two types of acceptances and rejections—

1. false rejections in which workers who failed the test were deemed to have acceptable job performance,
2. true acceptances, workers who scored well on the test and were acceptable on the job,
3. true rejections, workers who scored poorly on the test and had unacceptable job performance, and
4. false acceptances, workers who scored well on the test and had unacceptable job performance.



**Figure 6.4 Examples of contingency table**

The selection ratio, or percentage of time that one expects to be accurate, is calculated by summing the number of true acceptances and true rejections and dividing by the total number of examinees. That is—

$$\text{Correct Decisions} = \frac{(\text{True Acceptances} + \text{True Rejections})}{(\text{True Acceptances} + \text{True Rejections} + \text{False Acceptances} + \text{False Rejections})}$$

Using the data presented in Figure 6.4, the percentage of correct decisions would be 86.7 percent. With this approach, the accuracy of several potential test passing scores as predictors of performance for the data in Figure 6.4 can be investigated.

$$\text{Correct Decisions} = 0.867 = \frac{(18+8)}{(18+8+2+2)}$$

The previous example of an expectancy table for the tire builder job showed that either a score of 431.7 (10% level) or a score of 455.1 (20% level) provided substantial increases in job performance (Table 6.2). Table 6.3 shows that for individuals who achieve a test score of 431.7 or higher, 90% would have acceptable job performance (true acceptances) and 10% would not (false acceptances). For those individuals whose test score was below 431.7, 80% were deemed to have unacceptable job performance (true rejections) and 20% had acceptable job performance (false rejections). The accuracy of the pass/fail ratios or true acceptance and true rejections for the score of 455.1 are similar to the score of 431.7 (i.e., 91%, 72%, respectively). However, for a score of 483.8 (30% level) the true acceptance ratio (93%) remains high, but the true rejections drop dramatically (51%). This drop indicates that 49% of the individuals who scored below 483.8 and failed the test were deemed to have acceptable job performance. Therefore, use of a passing score of 483.8 would not be reasonable because it would eliminate many individuals who could perform the job. Setting the passing score at 455.1 or 431.7 would provide a more accurate assessment of job success.

**Table 6.3 Comparison of test score to acceptable and unacceptable job performance**

	Job Performance Level			
	Of those incumbents obtaining specific test score the expected job performance would be:		Of those incumbents failing at a specific test score the expected job performance would be:	
Minimum Test Passing Score	Unacceptable %	Acceptable %	Unacceptable %	Acceptable %
no test used	21	79	—	—
431.7 (10%)	10	90	20	80
455.1 (20%)	9	91	28	72
483.8 (30%)	7	93	49	51
518.7 (40%)	4	96	56	44
550.0 (50%)	2	98	66	34
569.5 (60%)	3	97	71	29
595.7 (70%)	4	96	76	24
605.9 (80%)	6	94	79	21
647.3 (90%)	11	89	84	18

**Combined Use of Expectancy and Contingency Tables**—To determine which score to use as the passing score, the expectancy table (Table 6.2) can be consulted to determine which score resulted in a greater increase in predicted job performance. Table 6.2 shows that the increase in job performance at the 10 percent level (431.7) is 6.1 points and the increase at the 20 percent level (455.1) is 7.2 points. The greatest increase in job performance of 7.8 point is found between the 20 percent (455.1) and 30 percent (483.8). However, the accuracy of true rejections (i.e., 51%) as shown in Table 6.3 is lower than the 20 percent level (i.e., 72%). Therefore, since the decision accuracy for the 10 percent (80 true rejections; 90% true acceptances) and 20 percent (72% true rejections; 91% true

acceptances) levels is comparable, the greatest utility would be achieved if the passing score were set at the 20 percent level or a passing score of 455.1 to obtain a greater expected job performance level.

## Normative Data

Scores from standardized or widely administered physical tests (e.g., sit-ups) can be used to compare one group to the original reference group. Such a comparison would determine an individual's status relative to his/her normative group. This type of testing and use of test scores, which uses previously gathered data for comparison, is called norm-referenced testing (Nitko, 1984; Safrit & Wood, 1989). When using norm-referenced testing, information about an individual's position relative to the normative group can be expressed in percentile ranks, normalized standard scores, or a descriptive classification such as "good" or "average." Although other scoring methods are used with norm-referenced tests, these are the most common encountered in physical testing.

Norm-referenced testing can be useful in determining an individual's ability relative to others in the same reference group. For example, a 30-year old woman who obtains a score of 18 inches on a flexibility test is considered "average," with a percentile rank of 45 when compared with other women of the same age (Golding, Myers, & Sinning, 1989). A 30-year old man with the same flexibility score would be considered "excellent" with a percentile rank of 75 when compared to men in the same age group. If the wrong reference group (e.g., 30-year old women) were used, this man would be misclassified. Use of the correct reference group is fundamental to classifying individuals accurately based on a testing procedure.

Most normative physical performance test data are collected on the "general population" by age and gender. Classifying individuals into above average, average, and fair groups may provide general information about an individual's relative standing. However, for employment purposes this classification method is inadequate. A passing score must represent the absolute level of performance needed for successful job performance. To merely identify a percentile level (e.g., 40%) and indicate that anyone who scores below this level fails the test is inadequate for making employment decisions, because it does not establish the relevance of the passing score (e.g., 40%) to future job performance.

To use the information obtained from a norm-referenced test in an employment setting, two criteria should be applied. First, the normative data must be drawn from a sample population representative of the prospective job population. It would be inappropriate to use a percentile rank (e.g., 40%) from normative general population data as the acceptable performance level to determine whether an applicant for a police officer position passes or fails the sit-ups test. The normative data for such a determination should be based on a normative police officer population. Second, the score (e.g., percentile) selected as the passing score must represent the level of performance needed for the job.

Finally, it should be noted that the Civil Rights Act of 1991 (Section 106) prohibits use of adjusted or different test passing scores based on race, gender, color, religion, or national origin. Use of different standards to compute norm-referenced scores for individuals in different groups (e.g., percentiles, T-scores) is unlawful in a selection setting. However, recent litigation (*Lanning v. SEPTA*, 1999) has challenged the use of a single passing score in favor of separate passing scores based on gender for tests that are assessing physical fitness.

## Ergonomic Data

Ergonomic or workplace data such as the energy cost of the work being performed, weights and dimensions of objects being lifted, distances objects are carried, and forces required to operate equipment have been used to establish passing scores for physical tests. Because passing scores should be set at a level that ensures minimally acceptable performance (Cascio, Outtz, Zedeck, & Goldstein, 1991; Cascio, Alexander, & Barrett, 1988), ergonomic data help define the minimal job demands. Ergonomists and physiologists have defined the energy costs (e.g., oxygen consumption, kilocalories per minute) and strength demands of work for job tasks ranging from light industrial to heavy tasks involved in firefighting, coal mining, and other manual materials handling jobs (Astrand & Rodahl, 1986; Ayoub, 1991). These ergonomic parameters have been used to match the worker to the job demands, as well as to modify the workplace design.

Research using ergonomic data to identify the energy costs of fighting fires and the forces required to lift patient-loaded gurneys into an ambulance has shown that job-related passing scores can be identified for aerobic capacity and strength tests. A study conducted using experienced firefighters indicated that a  $\text{VO}_2\text{max}$  of  $33.5 \text{ ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$  was needed to meet the demands of a sequence of essential firefighting tasks (Sothmann, Saupe, Jasenof, Blaney, Fuhrman, Woulfe, Raven, Pawelczyk, Dotson, Landy, Smith, & Davis, 1990). In this study, contingency tables were generated based on this value and were used to classify firefighters as acceptable and unacceptable. These contingency tables indicated that the true acceptances were 83%, and the true rejections were 67%. Lowering the acceptable  $\text{VO}_2\text{max}$  to  $30.5 \text{ ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$  resulted in only 25% of the firefighters being able to achieve acceptable job performance. Recent research using a criterion-related validity study compared incumbent firefighters' job performance with several strength and aerobic capacity measures and found that a  $\text{VO}_{2\text{submax}}$  of  $33.5 \text{ ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$  was indicative of acceptable job performance (Gebhardt, Baker, & Sheppard, 1995; Sothmann et al., 1995). Thus, a firefighter physical test that requires a  $\text{VO}_2\text{max}$  of  $33.5 \text{ ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$  would be appropriate for selection firefighters.

In another study, the minimum force required to lift the head-end of an ambulance gurney was computed using a biomechanical model. Ergonomic data such as the gurney weight, length, width, and height at full extension were obtained, along with the average weight and gender of the patients (Gebhardt, 1990). A model using these variables was designed to determine the force required to lift the head- and foot-ends of the gurney. The model indicated that the minimally acceptable force to lift the head-end of the patient-loaded gurney from the ground to full extension was 154.5 pounds. A criterion-related validity study conducted with incumbent paramedics used an instrument that simulated the gurney lifting position, along with other muscular strength and endurance tests. This study found that a force of 157.7 pounds on the simulated gurney lift was the minimum force associated with acceptable job performance. This force closely approximated the force value obtained in the biomechanical model.

Clearly, ergonomic data can be used effectively to define levels of performance. In addition to the examples above, other measures, such as production rates, can be used to identify a minimum level of performance. However, Jackson cautioned researchers that use of published energy cost data is problematic, in that it represents average values for the general population (Jackson, 1994). As

with norm-reference testing, energy costs across workers (e.g., by gender) may be equivalent, but an individual's actual work or power output may differ.

---

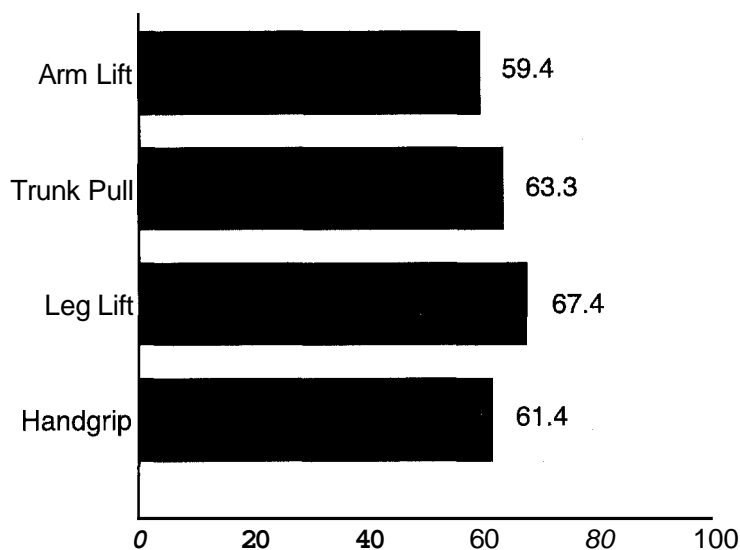
## Test Fairness and Adverse Impact

Adverse impact and test fairness go hand in hand when evaluating physical performance tests and establishing passing scores. The EEOC Uniform Guidelines (1978) define adverse impact by the four-fifths (4/5s) or 80% rule in which the passing rate for the minority group (e.g., women) is less than 80% of the pass rate for the majority group (e.g., men). A detailed discussion of the computation of adverse impact can be found in the Chapter 7 of this State of-the-Art Report (SOAR). Adverse impact in physical testing is based primarily on gender (Gebhardt, 2000; Hogan 1991). A test that had no adverse impact might be considered to be a fair test. However, Guion (1966) indicated that test fairness in an employment setting includes more than just lack of adverse impact. His definition of test fairness is "Individuals with equal probabilities of success have equal probabilities of being hired."

## Adverse Impact of Physical Tests

Past research has confirmed that men and women have significant strength differences (Gebhardt, 1999; Gebhardt et al., 1985, 1998b, 1999; Hogan, 1991; Jackson, 1994; McArdle, Katch, & Katch, 1996; Myers, Gebhardt, Crump, & Fleishman, 1983a; NIOSH, 1981). The greater the physical demands of a job, the less likely a subgroup of women will achieve an 80% pass rate of men. This is demonstrated by an example provided by Jackson (Jackson, 2000, Table 1) in which women and men are compared on an incremental lifting task that starts at 22 pounds and continues in 11-pound increments to 99 pounds. Using the 80% or 4/5s rule, adverse impact was shown at a lift of 55 pounds. These results are similar to a large U.S. Army study that used incremental lifting and found significant differences between men and women before and after basic training (Myers, Gebhardt, Crump, & Fleishman, 1983b). Since most physically demanding jobs require muscular strength, tests that evaluate the physical abilities needed for the job will have a muscular strength component. Therefore, it can be inferred that most physical tests will have an adverse impact on women as defined by the EEOC Uniform Guidelines.

The EEOC Uniform Guidelines also state that an alternative assessment with less adverse impact should be used. Many test developers believed that use of a work sample or job simulation test would reduce the adverse impact observed with basic ability tests (e.g., strength). However, research has shown that both have adverse impact. Figure 6.5 shows means for men and women across several basic ability strength tests. In all tests men perform significantly better than women, and the women's scores as a percentage of the men's ranged from 59% to 67% (Gebhardt, 1999). Similarly, Figure 6.6 illustrates that women's performance on job simulation tests as a proportion of men's performance is similar to that shown for basic ability tests in Figure 6.5. For example, the women's means for job simulation tests such as setting and climbing ladders, applying sidewall to



**Figure 6.5 Basic ability tests: Women's performance as percent of men's.** From "Applying regression analysis and biomechanical modeling to setting of cut-scores or qualifying standards," by D.L. Gebhardt, 1999, paper presented at American College of Sports Medicine, Seattle, WA. Reprinted with permission of author.

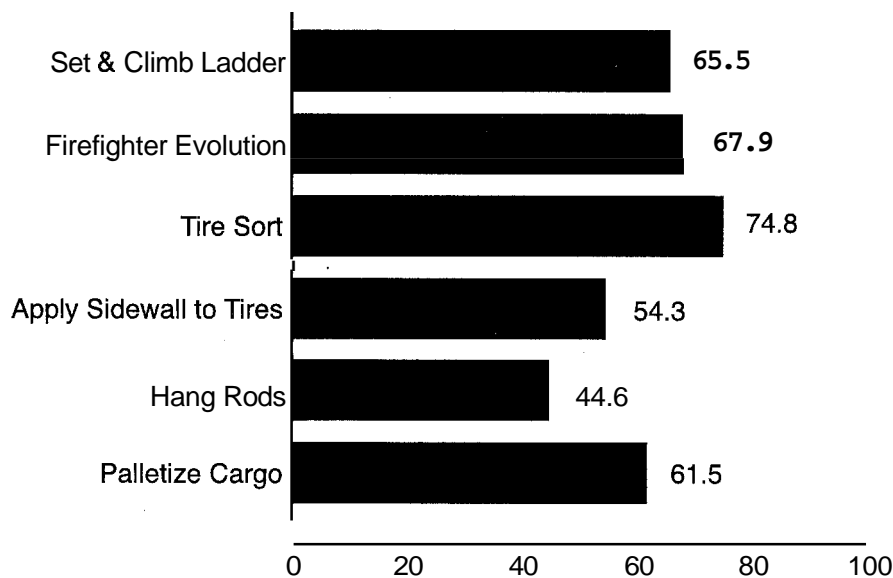
tires, and hanging 14-foot rods on containers, all of which required upper body strength, had similar or lower percentage proportions to the basic ability upper body strength tests (e.g., arm lift). Therefore, use of a job simulation may not be an effective alternative assessment procedure as defined by the EEOC Uniform Guidelines.

Finally, the magnitude of the difference between mean scores for men and women on physical tests is large. The statistical method to determine the size of these differences is called *effect size*, with a value of 0.8 or greater being defined as a large effect size (Cohen, 1988). The effect sizes found in a variety of studies ranged from .8 to 1.5 (e.g., Baker, Sheppard, Gebhardt, & Leonard, 2000; Gebhardt et al., 1998a; Marcinko, Nelson, Schneider, & Sproule, 1997). Therefore, the statistical difference between men's and women's physical test means is very large, as found in many studies (Gebhardt et al., 1998b, 1999, Reilly, Zedeck, & Tenopir, 1979; Sothmann et al., 1995).

## Test Fairness

To use a test to predict applicants' future job performance, the test must valid. Past research has shown that physical tests do indeed have high levels of validity (e.g., Gebhardt, 1985, 1998a & b, 1999; Jackson, 1994; Myers et al., 1983; Reilly et al., 1979; Sothmann et al., 1995). Even though an adverse effect impact on women was found in these studies (e.g., Gebhardt 1999; Sothmann et al., 1995), it does not necessarily follow that the tests were discriminatory or in violation of Federal statutes. These statutes state that tests with adverse impact may be used if the employer can demonstrate the job-relatedness and validity of the tests. Therefore, tests with adverse impact may be con-





**Figure 6.6 Job simulation tests: Women's performance as percent of men's.** From *"Applying regression analysis and biomechanical modeling to setting of cut-scores or qualifying standards,"* by D.L. Gebhardt, 1999, paper presented at American College of Sports Medicine, Seattle, WA. Reprinted with permission of author.

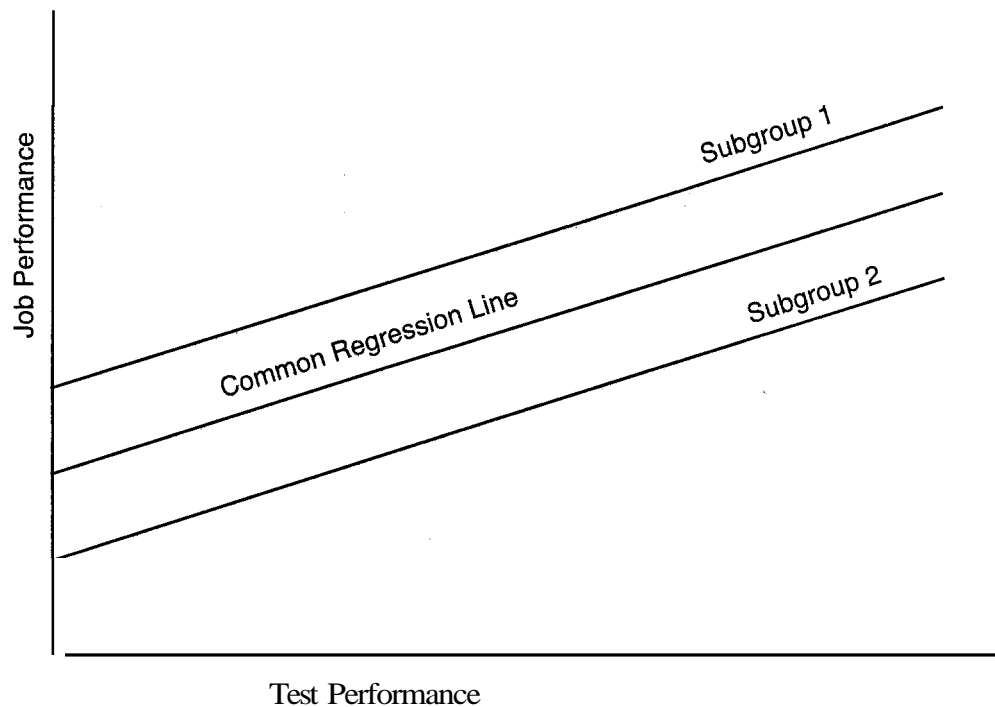
sidered fair. In many of the studies cited above, the tests had adverse impact on women but were found to be fair across subgroups (e.g., gender, race) because the tests measured specific abilities or behaviors needed for effective job performance. Thus, these tests are more likely to retain validity across different subgroups such as men and women (Anastasi, 1996).

To assess whether a test is valid for multiple groups (e.g., gender) or biased toward one subgroup, it is necessary to determine whether—

1. the validity coefficients or regression lines are similar across subgroups and
2. the test and/or criterion scores differ across subgroups (Bartlett, Bobko, Mosier, & Hannan, 1978; Kerlinger & Pedhazur, 1973).

The statistical procedure used for this analysis is differential prediction in which a moderated multiple regression analysis is employed to examine whether the validity coefficients (slope) and test and/or criterion mean scores (intercept) across subgroups differ from the overall regression equation (Bartlett et al., 1978; Kerlinger & Pedhazur, 1973). In the physical test domain, test fairness analyses are typically used to determine whether a test is biased in relation to gender, ethnic, and age groups.

Figure 6.7 illustrates the slope and intercept for the common regression line generated using the total sample population, along with the regression lines for example subgroups 1 and 2. If the Y-intercept of each subgroup regression line differs from the Y-intercept of the common regression line, it would be concluded that the subgroups differed on the physical tests used as predictors and/or on the criterion measure. In the physical testing arena, the regression line for subgroup 2 is usually represented by women's performance, with subgroup 1 representing the men. Studies that



**Figure 6.7** *Differential prediction analysis*

investigated test fairness for physical tests have found both significant and non-significant intercept differences for gender (Baker et al., 2000; Gebhardt, 1985, 1998b; Reilly et al., 1979).

The slopes of the subgroup regression lines are also tested for differences between each subgroup and the common regression line. If significant differences in the slopes are found, this may be attributed to variations in a subgroup's physical test score and/or criterion measure score. This finding is more problematic than intercept differences because the differential group validity indicates that separate regression equations are appropriate for each group. However, the Civil Rights Act of 1991 prohibits adjusting test scores based on subgroup differences.

Test fairness analysis can help to increase test utility in relation to adverse impact. Adverse impact can be reduced or eliminated by lowering a passing score. However, in the physical domain, lowering the passing score to a point that eliminates gender adverse impact typically diminishes the utility of the test. Physical tests that have utility usually have adverse impact. Although adverse impact is undesirable, the Federal statutes indicate that it is acceptable as long as the test is valid, job-related, and fair. Use of the test fairness analysis satisfies one of these conditions.

---

## Summary

In this chapter several methods for determining passing scores were discussed. Use of a specific technique (i.e., expectancy tables, contingency tables, Taylor Russell tables, normative data,

ergonomic data) depends upon the type of model used to validate the test (i.e., content, criterion-related, construct). The need for accurate test (predictor), job performance (criterion), and ergonomic data was emphasized. The integration of data from multiple sources, such as expectancy and contingency tables, is important in establishing an accurate passing score that will reflect future job performance. Finally, factors such as test fairness and adverse impact must be considered when setting a passing score to ensure accurate employment decisions and compliance with the EEOC Uniform Guidelines (1978).

---

## References

1. American College of Sports Medicine. (2000). *ACSM's guidelines for exercise testing and prescription* (6<sup>th</sup> ed.). Philadelphia, PA: Lippincott, Williams, & Wilkins.
2. American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
3. Anastasi, A. (1996). *Psychological testing*. Englewood Cliffs, NJ: Prentice Hall.
4. Astrand, P. O., & Rodahl, K. (1986). *Textbook of work physiology*. New York: McGraw-Hill, Inc.
5. Ayoub, M. A. (1991). Determining permissible lifting loads: An approach. *Proceedings of Human Factors Society 35<sup>th</sup> Annual Meeting*, 35, 825–829.
6. Bartlett, C. J., Bobko, P., Mosier, S. B., & Hannan, R. (1978). Testing for fairness with a moderated multiple regression strategy: An alternative to differential analysis. *Personnel Psychology*, 31, 233–241.
7. Baker, T. A., Sheppard, V. A., Gebhardt, D. L., & Leonard, K. M. (2000). *Development and validation of physical performance test for selection of City of Lubbock police officers*. Beltsville, MD: Human Performance Systems, Inc.
8. Casio, W. F., Alexander, R. A., & Barrett, G. V. (1988). Setting cutoff scores: Legal, psychometric, and professional issues and guidelines. *Personnel Psychology*, 41, 1–24.
9. Casio, W. F., Outtz, J., Zedeck, S., & Goldstein, I. L. (1991). Statistical implications of six methods of test score use in personnel selection. *Human Performance*, 4, 233–264.
10. Cederbloom, D. (1989). Peer and supervisor evaluation: An underused promotion method used for law enforcement. *Proceedings of 13<sup>th</sup> Annual Meeting of International Personnel Management Association-Assessment Council*.
11. Civil Rights Act of 1991, S. 1745, 102<sup>nd</sup> Congress, (1991).
12. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
13. Crump, C. E., Gebhardt, D. L., Guerette, P. J., & Wertheimer, N. M. (1985). *Development and validation of entry-level physical performance tests for Uniroyal tire plants*. Bethesda, MD: Advanced Research Resources Organization.
14. Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, and Department of Justice. (1978). *Uniform Guidelines on Employee Selection Procedures*. Washington, DC: Bureau of National Affairs, Inc.
15. Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51, 327–358.

16. Gebhardt, D. L. (1999) *Applying regression analysis and biomechanical modeling to setting cut-scores or qualifying standards*. Paper presented at the American College of Sports Medicine, Seattle, WA.
17. Gebhardt, D. L. (1990). *Validation of the force required to lift a patient-loaded gurney*. Paper presented at the American College of Sports Medicine, Salt Lake City, UT.
18. Gebhardt, D. L., Baker, T. A., & Sheppard, V. A. (1998a). *Volume 2: Q-2 police officer physical performance test and medical guidelines development and validation report*. Hyattsville, MD: Human Performance Systems, Inc.
19. Gebhardt, D. L., Baker, T. A., & Sheppard, V. A. (1998b). *Development and validation of physical performance tests for the selection and fitness assessment for United States Postal Inspection Service candidate and incumbent Postal Inspectors*. Hyattsville, MD: Human Performance Systems, Inc.
20. Gebhardt, D. L., Baker, T. A., & Sheppard, V. A. (1995). *Development and validation of physical performance tests for City of Chicago Firefighters. Volume 2: Job analysis*. Hyattsville, MD: Human Performance Systems, Inc.
21. Gebhardt, D. L., Baker, T. A., Sheppard, V. A., & de Miranda, G. A. (1994). *Development and validation of physical performance tests for emergency medical technicians and paramedics*. Hyattsville, MD: Human Performance Systems, Inc.
22. Gebhardt, D. L., Baker, T. A., Sheppard, V. A., & Leonard, K. M. (1999). *Development and validation of physical performance tests for selection of City of Chicago paramedics*. Beltsville, MD: Human Performance Systems, Inc.
23. Gebhardt, D. L., Schemmer, F. M., & Crump, C. E. (1985). *Development and validation of selection tests for longshoremen and marine clerks*. Bethesda, MD: Advanced Research Resources Organization.
24. Giffin, M. E. (1989). Personnel research on testing, selection and performance appraisal. *Public Personnel Management*, 18, 127–137.
25. Golding, L. A., Myers, C. R. & Sinning, W. E. (1989). *Y's way to physical fitness*. Champaign, IL: Human Kinetics Publishers, Inc.
26. Guion, R. M. (1966). Employment tests and discriminatory hiring. *Industrial Relations*, 5, 20–37.
27. Hogan, J. C. (1991). The structure of physical performance in occupational tasks. *Journal of Applied Psychology*, 76, 495–507.
28. Jackson, A. S. (2000). Legal issues. In S. Constable & B. Palmer (Eds.). *The Process of physical standards development — State of the art report*. (HSIAC-SOAR-2000-001) Wright-Patterson AFB, OH: ??????
29. Jackson, A. S. (1994). Preemployment physical evaluation. In J. O. Holloszy (Ed.), *Exercise and sport sciences reviews* (Volume 22). Baltimore, MD: Williams & Wilkins.
30. Kerlinger, F. N., & Pedhazur, E. J. (1973). *Multiple regression in behavioral research*. New York, NY: Holt, Rinehart & Winston, Inc.
31. King, L. M., Hunter, J. E., & Schmidt, F. L. (1980). Halo in a multidimensional force-choice performance evaluation scale. *Journal of Applied Psychology*, 65, 507–516.
32. Landy, F. J. & Farr, J. L. (1983). *The measurement of work performance: Methods, theory, and applications*. New York: Academic Press.
33. Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin*, 87, 72–107.
34. Lanning v. Southeastern Pennsylvania Transportation Authority, LEXUS 14607, (3<sup>rd</sup> Cir., 1999 U.S. App).

35. Latham, G. P., & Wexley, K. N. (1977). Behavioral observation scales for performance appraisal purposes. *Personnel Psychology*, 30, 225–268.
36. Marcinko, M. R., Nelson, E. R., Schneider, R. E. & Sproule, C. F. (1997). *Corrections officer trainee and corrections officer I physical/medical standards project: Validation report*. Harrisburg, PA: Commonwealth of Pennsylvania State Civil Service Commission and Department of Corrections.
37. McArdle, W. D., Katch, F. I. & Katch, V. L. (1996). *Exercise physiology*. Malvern, P A Lea & Febiger.
38. Myers, D. C., Gebhardt, D. L., Crump, C. E., & Fleishman, E. A. (1983a). The dimensions of human physical performance: Factor analysis of strength, stamina, flexibility, and body composition measures. *Human Performance*, 6, (4), 309–344.
39. Myers, D. C., Gebhardt, D. L., Crump, C. E., & Fleishman, E. A. (1983b). *Validation of the military enlistment physical strength capacity test*. Bethesda, MD: Advanced Research Resources Organization.
40. Mumford, M. D. (1983). Social comparison theory and the evaluation of peer evaluations: A review and some applied implications. *Personnel Psychology*, 36, 867–881.
41. Naylor, J. C., & Shine, L. C. (1965). A table for determining the increase in mean criterion score obtained by using a selection device. *Journal of Industrial Psychology*, 3, 33–42.
42. Nitko, A. J. (1984). Defining "criterion-referenced tests." In R. A. Berk (Ed.). *A guide to criterion-referenced test construction*. Baltimore, MD: Johns Hopkins University Press.
43. NIOSH. (1981). *Work practices guide for manual lifting*. NIOSH Technical Report No. 81–122. Cincinnati, OH: National Institute for Occupational Safety and Health and U.S. Department of Health and Human Services.
44. Reilly, R. R., Zedeck, S., & Tenopir, M. L. (1979). Validity and fairness of physical ability tests for predicting craft jobs. *Journal of Applied Psychology*, 64, 267–274.
45. Safrit, M. J., & Wood, T. M. (1989) *Measurement concepts in physical education and exercise science*. Champaign, IL: Human Kinetics Books.
46. Senate Report No. 101–116, 101<sup>st</sup> Congress, 1<sup>st</sup> Session, 21; House Report 101–485, Part 2, 101<sup>st</sup> Congress, 2<sup>nd</sup> Session, 51 (1990).
47. Smith, P. C., & Kendall, L. M. (1963). Retranslation of expectation: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology*, 47, 149–155.
48. Sothmann, M. S., Gebhardt, D. L., Baker, T. A., Costello, G., & Sheppard, V. A. (1995). *Development and validation of physical performance tests for City of Chicago Firefighters. Volume 3 Validation of physical tests*. Hyattsville, MD: Human Performance Systems, Inc.
49. Sothmann, M. S., Saupe, K., Jasenof, D., Blaney, J., Fuhrman, S., Woulfe, T., Raven, P., Pawelczyk, J., Dotson, C., Landy, F. Smith, J., & Davis, P. (1990). Advancing age and the cardiorespiratory stress of fire suppression: Determining a minimum standard for aerobic fitness. *Human Performance*, 3, 217–236.
50. Taylor, H. C., & Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. *Journal of Applied Psychology*, 23, 565–578.



# Chapter 7

## Legal Issues

Andrew S. Jackson, P.E.D.

Department of Health and Human Performance

University of Houston

Houston, TX

---

### Abstract

This chapter is an overview of the legal forces and issues related to employment practices. Title VII of the Civil Rights Act of 1964, the Age Discrimination in Employment Act (ADEA) of 1967, and Americans with Disabilities Act (ADA) of 1990 are the federal laws that define discriminatory employment practices. The centerpiece of employment discrimination law is Title VII of the Civil Rights Act of 1964, as amended by Congress on several occasions. Title VII prohibits employment discrimination because of “race, color, religion, sex, and national origin” by employers, labor organizations, and employment agencies. Title VII tends to be comprehensive in that everyone is potentially covered, because both genders and all majority and minority racial and ethnic groups, as well as religious groups, are covered by Title VII, but the act does not apply to Military personnel.

The disparate impact theory is used to establish employment discrimination. This legal process has a three-part burden of proof. First, the plaintiff (employee) must establish that the hiring practice has a disparate impact on a protected group. Although not legally mandated, the Equal Employment Opportunity Commission (EEOC) Guidelines are often used to define disparate impact. The guidelines use the four-fifths (4/5s) rule to define adverse impact. Under the 4/5s rule a selection device has adverse impact when the pass rate for one protected group is less than four-fifths, or 80 percent, of the pass rate of the group with the highest pass rate. Once adverse impact is established, the burden of proof then falls on the defendant (employer) to justify that the exclusionary effect is a business necessity. The defendant must show that the selection method is job related. This involves demonstrating that the selection device (e.g., preemployment test) is valid. A common method used to establish job relatedness is with a validation study. Lastly, if business necessity is established, the burden of proof shifts back to the plaintiff to demonstrate that the employer failed to use a selection device that is equally effective but has a lesser disparate impact.

This chapter reviews cases related to physical testing. Many of these cases involve the use of height and weight standards and tests for selecting public service employees such as police officers and firefighters. The outcome of this litigation largely depends on the scientific quality of validation study. The recent court ruling of *Lanning v. SEPTA* (U.S. 3rd Circuit 1999) will likely have a major impact on physical testing. An aerobic fitness cut-score representing a  $\text{VO}_2\text{max}$  of 42.5 ml/kg/min was found to be unacceptable by the court. An option offered by the court was the validation of an aerobic fitness cutoff score that measures the minimum capacity necessary to perform the job. This court ruling is consistent with established physiological and ergonomic principles of selecting workers with the fitness demanded by the job. This ruling suggests that validation studies will be evaluated not only by standard psychometric criteria but also by physiological validation of the test and cut-score.

---

## Introduction

Employers have always used some method to select an employee from potential job applicants. The rapid rise in the use of standardized tests for job placement can be traced to this country's need for rapid mobilization and use of human resources during the first and second World Wars. The goal was to match Military personnel to jobs on the basis of test performance. The development of pre-employment tests grew out of the discipline of psychology and their early success in measuring differences among people. The common theme of this work was that persons differ from each other, in reasonable stable ways, on some number of attributes, and that patterns of individual attributes are more or less suited to particular patterns of job requirements (Dunnette & Hough, 1991).

Much of the early preemployment testing focused on cognitive abilities, but with the rise in women seeking jobs that were once male dominated, the need for preemployment physical abilities tests increased. The need for valid tests for selecting personnel for physically demanding jobs can be traced to at least three important forces. First, equal employment opportunity legislation resulted in greater numbers of women and handicapped persons seeking employment in occupations requiring high levels of physical ability. Second, evidence suggested that physically unfit workers had higher incidences of lower back injuries. Lastly, preemployment medical evaluations used alone are inadequate for selecting personnel for physically demanding jobs. The disciplines most prominent in physical ability employment testing are industrial-organizational (I/O) psychology, industrial engineering, ergonomics, biomechanics, and exercise physiology.

This section provides an overview of employment law as it relates to defining job discrimination in the civilian sector of the United States. This section also provides a brief overview of the Federal laws used to define job discrimination. This is followed by the legal process used to describe discrimination. Next, cases relevant to physical testing and exercise physiology are examined. Lastly and quite hazardously, possible future legal directions are explored.



---

## Legal and Legislative Forces

The current interest and research on preemployment test methodology for physically demanding jobs has its roots not only in work physiology (Astrand & Rodahl, 1970; Durnin & Passmore, 1967; McArdle, Katch, & Katch, 1991; Passmore & Durnin, 1955), and psychometric test theory (Division of Industrial-Organizational Psychology & Association, 1987), but also in Federal civil rights legislation and court decisions on employment practices. Title VII of the Civil Rights Act of 1964, the Age Discrimination in Employment Act (ADEA) of 1967, and Americans with Disabilities Act (ADA) of 1990 are the Federal laws used for employment litigation. Although Title VII and ADEA tend to be unambiguous, ADA has been found to be more difficult to interpret.

The centerpiece of employment discrimination law is Title VII of the Civil Rights Act of 1964, as amended by Congress on several occasions. Title VII prohibits employment discrimination on the basis of “race, color, religion, sex, and national origin” by employers, labor organizations, and employment agencies. The term “sex” refers to “gender” and does not include sexual orientation (Rothstein, Craver, Shroeder, & Shoben, 1999). Title VII tends to be comprehensive in that everyone is potentially covered—both genders and all majority and minority racial and ethnic groups, as well as religious groups, are covered by Title VII. The act does not apply to Military personnel (Rothstein et al., 1999).

The Age Discrimination in Employment Act (ADEA) in 1967 provides the legal basis for defining job discrimination on the basis of age. The substantive provisions of the ADEA Act are identical to Title VII with the substitution of the word “age” as the prohibited basis for discrimination in place of “race, color, religion, sex and national origin” found in Title VII (Rothstein et al., 1999, p. 215).

The most recent law used to define discrimination is the American with Disabilities Act (ADA) of 1990. ADA is a comprehensive federal law that prohibits discrimination in a wide variety of segments of life. The law has five titles. Title I covers employment of Americans with physical and mental disabilities (Rothstein et al., 1999). The law defines a person with disabilities as someone with a substantial impairment that significantly limits or restricts a major life activity such as hearing, seeing, speaking, walking, breathing, performing manual tasks, caring for oneself, learning, or working.

According to section 101(8) of ADA, a disabled worker is a person with a disability who, with or without reasonable accommodation, can perform the essential functions of the employment position (Rothstein et al., 1999). Employers have a duty to make reasonable accommodations to the known physical or mental disability. Some examples are making facilities accessible, job restructuring, acquisition or modification of equipment or devices. Reasonable accommodation is not required if it results in undue hardship to the employer defined as “an action requiring significant difficulty or expense in light of factors such as the nature and cost of the accommodation and the size and financial resources of the company” (Rothstein et al., 1999, p. 246). The exact number of Americans covered under this law is not known, but it has been estimated to exceed 43 million.

A key issue with the ADA is that the person must be legally disabled. What substantially limits a major life activity, the legal definition of a disability, is currently defined in the courts. In 1999 the Supreme Court ruled in two cases that having a condition that is correctable is not considered a disability. In the case of *Sutton v. United Airlines, Inc.* [\_\_ U.S. \_\_, 119 S.Ct. 2139, 144 L.Ed.2d 450 (1999)] the effect of eyeglasses on vision-impaired plaintiffs should be considered when defin-

ing a disability under the ADA. In the companion case of *Murphy v. United Parcel Service, Inc.*, the Court ruled that in evaluating the severity of the plaintiff's hypertension, the effect of his medication should be considered [\_\_ U.S. \_\_, 119 S.Ct. 2133, 144 L.Ed.2d 450 (1999)]. These two rulings showed that the burden of documenting a disability clearly rests with the employee and that an employee who fails to control a controllable disability may lose his or her protection under ADA (Rothstein et al., 1999).

Under ADA, preemployment medical examinations and medical inquiries are illegal. Although preemployment medical examinations may not be given, a post-offer medical examination is permitted and a conditional job offer may be withdrawn if the examination documents that the applicant is unable to perform the essential functions of the job. Medical qualification for a job is a two-step process. First, the physical and mental demands of the job must be documented. Second, the medical examination must evaluate the applicant's capacity to perform the essential job functions. In addition, medical examinations can be used to determine if a person is physically able to return to work after a disability leave (Rothstein et al., 1999).

In contrast to a general medical examination, a preemployment skill or physical ability test can be legally used for employee selection. Under ADA such a test is not considered a medical examination. However, if the preemployment test screens out applicants on the basis of disability defined under ADA, the employer has the burden of proving that the test is job related and consistent with business necessity (Rothstein et al., 1999).

---

## Legal Process—Discrimination Litigation

The disparate impact theory is used to establish discrimination under Title VII, ADEA and ADA. This legal process has a three-part burden of proof—

1. The plaintiff (employee) must establish a disparate impact on a protected group.
2. If disparate impact on a protected group is established, the defendant (employer) must then justify the exclusionary effect with a business necessity. The defendant must show that the selection method is job related.
3. If business necessity is established, the burden of proof shifts back to the plaintiff to demonstrate that the employer failed to use a selection device that is equally effective but has a lesser disparate impact.

### Disparate Impact

In order to find an employment selection method discriminatory, the plaintiff (i.e., the job applicant or affected employee) must establish that the method has disparate impact on a protected group. The plaintiff must show that the employment selection method adversely affects the employment opportunities based on race, color, religion, sex, national origin, age, or a qualified disability. The Supreme Court case *Griggs v. Duke Power Co.* (401 U.S. 424, 91 S.Ct. 849, 28) was

the first case to use a disparate impact theory of discrimination. The power company used standardized aptitude tests for assigning jobs. The plaintiff class showed that the aptitude tests adversely affected racial groups. While 58 percent of whites passed the test, only 6 percent of African-Americans passed. The courts ruled that under Title VII, employment tests with disparate impact could not be used unless they were job related (Rothstein et al., 1999).

Although disparate impact or adverse impact must be proved to have discrimination, the Federal laws do not define explicitly what constitutes it. In 1966 the Equal Employment Opportunity Commission (EEOC) published the first set of guidelines on employment testing that were revised in 1970. This led in 1978 to the publication of the *Uniform Guidelines on Employee Selection Procedure!* (EEOC, 1991). These Federal standards and rules were jointly agreed on by the EEOC, Civil Service Commission, and Departments of Labor and Justice. The EEOC Guidelines use the four-fifths ( $4/5$ s) rule to define adverse impact. This tends to be a rule of thumb used by the EEOC and Federal enforcers of employment law to define adverse impact. Under the  $4/5$ s rule a selection device has adverse impact when the pass rate for one protected group is less than four-fifths, or 80 percent, of the pass rate of the group with the highest pass rate. To illustrate, assume the pass rate for the highest group is 60 percent and the pass rate of a protected group is 30 percent. In this example, the pass rate of the protected group is 50 percent ( $30\% \div 60\% = 50\%$ ) of the highest group. Under the EEOC Guidelines this would constitute disparate impact because it is below the 80 percent standard. If the pass rate of the protected group was 41 percent, while the pass rate of the highest group was 50 percent, the selection device would not have adverse impact. The pass rate of the protected group would be 82 percent ( $41\% \div 50\% = 82\%$ ), above the 80 percent standard defined in the EEOC Guidelines. Rothstein (Rothstein, 1999) points out that trial courts need not adhere to the  $4/5$ s rule, but legal history shows that the  $4/5$ s rule is viewed favorably by the courts.

When physical tests are used for employment decisions, sex tends to be a source of adverse impact (Hogan & Quigley, 1986; Hogan, 1991). This potential for adverse impact can be traced to the well-documented male and female differences in strength (Baumgartner & Jackson, 1999; Golding, Meyers, & Sinning, 1989; Hoffman, Stouffer, & Jackson, 1979; Laubach, 1976; NIOSH, 1977), maximal oxygen uptake ( $VO_{2max}$ ) (Astrand & Rodahl, 1970; Golding, Meyers, & Sinning, 1989; Jackson, Beard, Wier, Ross, Stuteville & Blair et al., 1995; Jackson, Wier, Ayers, Beard, Ross, Stuteville, & Blair et al., 1996; Vogel, Patton, Mello, & Daniels, 1986), and percent body fat (Jackson & Pollock, 1978; Jackson, Pollock, & Ward, 1980; McArdle, Katch, & Katch, 1991; Vogel et al., 1986; Wilmore & Costill, 1994).

Although much of the reason for adverse impact in physical testing can be traced to physiological differences between men and women, another factor is the physical demands of the job. The more physically demanding the job, the more likely a test will fail the EEOC Guidelines  $4/5$ s rule. Table 7.1 illustrates this with lifting data obtained on 608 women and men in the Human Factors lab at the University of Houston. The lift task was the common floor-to-knuckle height lift that became progressively heavier (Jackson & Sekula, 1999). The goal was to continue to lift heavier loads until the weight became too heavy to lift. Table 7.1 gives the percentages of men and women who could lift loads between 15 and 45 kg (33 to 99 pounds). The data in Table 7.1 illustrate that there would not be adverse impact based on the EEOC Guidelines  $4/5$ s rule for lift loads of  $< 25$  kg, but would have adverse impact for lift loads  $> 30$  kg. Further data in Table 7.1 show that as the physical demands of the task increase, the more likely it is that the task would have legally disparate impact.

Table 7.1 Male and female pass rates for floor-to-knuckle height lifts and the application of the 4/5s rule

Lift Weight kg (lbs)	Female Pass Rate (%)	Male Pass Rate (%)	4/5s Rule (%)
15 (22)	100.0	100.0	100.0
20 (33)	100.0	100.0	100.0
25 (44)	93.2	100.0	93.2
30 (55)	72.3	100.0	72.3
35 (66)	38.4	99.2	38.7
40 (77)	20.9	96.9	21.6
45 (99)	10.5	87.4	12.0

Business Necessity—Job Related

Once a plaintiff has established that a selection device has adverse impact (i.e., established a prima facie case), the defendant (employer) must prove that the test or selection device is job related (Rothstein et al., 1999). Just because a test is found to have adverse impact does not mean that the test or selection method is discriminatory. Once an employer becomes aware of the adverse impact, it becomes his or her responsibility to establish the validity of the requirement. Any adverse effect on the basis of race, gender, national origin, age, or disability is irrelevant as long as the selection device is valid (Rothstein et al., 1999). The Supreme Court case *Duke Power Co.* (401 U.S. 424, 91 S.Ct. 849, 28) was the first case not only to use a disparate impact theory of discrimination, but also to introduce the concepts of business necessity and job validation as defenses. "...Title VII forbids the use of employment tests that are exclusionary in effect unless the employer demonstrates that any given requirement has a "manifest relationship" to the job." (Rothstein et al., 1999, p. 163). The power company lost the case because the defendant could not show that an aptitude test found to have disparate impact against African-American employees was job related. The EEOC Guidelines provide valid methods of defining the job relatedness of a test or selection method and are used by the EEOC to evaluate hiring practices to determine that it is job related. Although the EEOC Guidelines' validation requirements have had a significant influence in Title VII litigation, the Supreme Court ruled that Title VII does not necessarily require employers to introduce formal validation studies showing that particular criteria predict actual on-the-job performance [Albemarle Paper Co. v. Moody, 422 U.S. 405, 95 S.Ct. 2362, 45 L.Ed.2d. 280 (1975)]. The legal obligation of employers who use an employment method with adverse impact is to prove its job relatedness. The Supreme Court decision of *Griggs v. Duke Power Co.* (401, U.S. 424, 1971) established that the employer must demonstrate that the selection method was job related and a business necessity (Rothstein et al., 1999). The position of the Court is that discriminatory tests cannot be used unless shown to be valid by professionally acceptable methods. In the case of *Albemarle Paper Co. v. Moody* (422 U.S. 405, 1975) the Court used the test validation methods outlined in the EEOC Guidelines to evaluate the discriminatory tests used by the paper company

to promote workers to more skilled positions and ruled that the validation studies were inadequate in several respects under the EEOC Guidelines (Rothstein et al., 1999).

Although the law does not require an employer to follow the procedures outlined in the EEOC Guidelines to establish job relatedness, the failure to do so encourages litigation. The EEOC Guidelines lists three acceptable types of validity studies. These are as follows—

1. Criterion-related validity is established with empirical evidence (e.g., correlation coefficients) linking a test and job. The goal is to demonstrate that the test predicts important elements of work behavior.
2. Content validity is established by replicating major portions of the job. The goal is to develop a test that is a representative sample of the behaviors of the job.
3. Construct validity demonstrates that a test measures identifiable traits or characteristics important for successful job performance. The goal is to show that the test accurately measures the construct and that the construct is necessary for successful job performance.

The EEOC Guidelines require that not only must the test be validated, but also the cutoff score must be validated. To illustrate, *Harless v. Duck* was a class action that challenged the physical ability test used by the Toledo, Ohio, Police Department. Applicants were required to complete three of the following four tests: 15 push-ups; 25 sit-ups; 6-foot standing broad jump; and a 25-second obstacle course. The Sixth Circuit endorsed the need for fitness but concluded that there is no justification in the record for the type of exercises chosen or the passing marks for each exercise. (Rothstein et al., 1999) In a 1999 case (*Lanning v. SEPTA*) argued at the U.S. Court of Appeals (3<sup>rd</sup> Circuit) a  $\text{VO}_2\text{max}$  cut-score of 42.5 ml/kg/min was found to be discriminatory under a disparate impact theory of liability for selecting police officers for a regional mass transit authority. Although evidence was presented showing that  $\text{VO}_2\text{max}$  was significantly correlated with job performance measured by arrest records, the Court ruled that the cut-score was discriminatory under the Civil Rights Act of 1991 because the validation study did not establish that this was the minimum qualification necessary for successful performance of the job.

## Alternative Selection

If the plaintiff establishes that a test or selection method has disparate impact, but the defendant proves the validity of the method, the plaintiff can still prove discrimination under the disparate impact theory of discrimination by demonstrating that there was a less discriminatory alternative. The less discriminatory alternative was established in *Albemarle Paper Co. v. Moody* case (422 U.S. at 405, 425, 95 S.Ct. At). The Court's explanation was as follows—

*If an employer does then meet the burden of proving that its tests are "job related," it remains open to the complaining party to show that other tests or selection devices, without a similarly undesirable racial effect, would also serve the employer's legitimate interest in "efficient and trustworthy workmanship." Such a showing would be evidence that the employer was using its tests merely as a "pretext" for discrimination.*

Less discriminatory alternatives not only pertain to the testing device, but also to setting the cut-score. Although *Albemarle Paper Co. v. Moody* case clearly established the burden of proof for the alternate selection prong of the disparate impact litigation, Rothstein et al. (1999) report that this prong is legally unclear and that case law has not provided much clarification. The Fifth Circuit Court considered the concept in *Brunet v. City of Columbus* (58 F.3d 251 5th Cir. 1995). The court rejected the plaintiff's claim that the city should have used a different cut-score for a firefighter test to have less adverse impact on female firefighters. The finding of fact in the case was that "there was not a substantially equally valid cutoff score with a lesser adverse impact, and it was not required to consider all possible alternative hiring procedures..." (Rothstein, 1999, p. 174). Similarly, in *Smith v. City of Des Moines* (99 F.3d 1466 8th Cir. 1996), the Court rejected the age discrimination claim and ruled that "the plaintiff failed to demonstrate that the proposed alternative would actually have a lesser impact nor that it would serve the city's legitimate interests equally well" (Rothstein et al., 1999, p. 174).

---

## Court Rulings

Arvey and Faley (1988) maintain that the landmark case of *Myart v. Motorola* in 1963 was the first signal that the court system became involved in the employment process. Leon Myart, a African-American with previous job-related experience, was refused a job in a Motorola plant because his score on a five-minute intelligence test was too low. Myart filed a complaint with the Illinois Fair Employment Practices Commission and charged racial discrimination. The Illinois Commission ruled that Myart be offered a job and ruled that the test could no longer be used for selection decisions. This landmark case motivated employers to develop preemployment tests that did not discriminate against a protected group.

Although the focus of this document is on preemployment testing, it is important to realize that most employment legal cases are for dismissals and failures to be promoted, not failure to hire. Rothstein and associates (Rothstein et al., 1999) point out that those who have been "wronged" on the job are more likely to initiate litigation than those who have failed to be hired. They further explain that the legal system is less sympathetic to those who have failed to be hired than those who lose their jobs.

Hogan and Quigley (1986) provide an excellent review of court cases related to physical testing and physical standards. The cases reviewed include the use of height and weight standards and physical ability testing. Of the 44 cases reviewed, 34 involved height and weight standards, and 10 involved physical ability tests for employee selection. Of these cases, most (37) involved law enforcement and firefighter employee-selection procedures. This section provides cases related to the following—

1. The use of height and weight
2. The use of physical tests for selecting firefighters
3. Cases related to lifting and materials handling
4. Cases related to the use of physiological parameters to define cut-scores.

## Height and Weight

In the 1960s, height and weight standards were a condition of employment for many public safety jobs, and these standards clearly had an adverse impact on women. Arvey and Faley (1988) reported that in 1973, nearly all of the nation's large police departments had a minimum height requirement. The average requirement was 68 inches. More than 90 percent of the women and 45 percent of the men would be expected to fail the 68-inch height requirement. The rationale for the standard was that size was related to physical strength, and the effectiveness of a police officer's job performance depended on strength.

An important case on the use of a height and weight requirement was decided in June 1977 by the U.S. Supreme Court. In *Dothard v. Rawlinson*, a woman was refused employment as a correctional-counselor trainee because she did not meet the minimum height and weight requirements of 62 inches and 120 pounds. The standard was found to have adverse impact because it excluded 33.3 percent of the women and only 1.3 percent of the men. The pass rate of females was only 67 percent of males, thereby failing the 4/5s rule and establishing adverse impact. Once adverse impact was established, the defendants argued that the height and weight requirements were job related because they have a relationship to strength, which is job related. The Supreme Court ruled that if strength is a real job requirement, then a direct measure of strength should have been adopted. The defendants failed to prove that the height and weight requirement was a business necessity.

The following cases provide instances in which height or weight requirements have been supported in the courts. In *Boyd v. Ozark Airlines*, the Court ruled in favor of a height requirement. The height requirement had disparate impact against women, but the Court ruled that a minimum height is necessary for a pilot to see properly and reach all the controls in an airplane cockpit (Hogan & Quigley, 1986). In *Costa (Le Boeuf) v. Markey* (DC, VA 1977) the Court ruled that imposing a 5'6" height requirement on a list of exclusively female applicants did not produce a disparate effect on women. The ruling of the Court in *EEOC v. Delta Air Lines* (DC, TX 1980) was that requiring different height and weight standards for male and female airline flight attendants did not constitute a sex discrimination violation of Title VII because weight is generally subject to one's own control. However, in another airlines case that Court ruled that it was sex discrimination to require female airline flight attendants to maintain certain weight levels and to have their weight monitored when similar requirements were not placed on male flight attendants (*Gerdorn v. Continental Airlines*, 9<sup>th</sup> Cir. 1982). In *Meadows v. Ford Motor Co.* (DC, KY 1973) the use of a minimum weight requirement for employees was unlawful discrimination on the basis of sex because the weight requirement was not shown to be related to job qualifications.

## Physical Tests

Of the 10 cases involving physical tests reviewed by Hogan and Quigley (1986), all involved police and firefighter preemployment tests. One of the cases (*Hull v. Cason*) was for race discrimination while the remaining charged sex discrimination under Title VII. The common test development approach that emerged from these cases was the use of general physical ability and fitness

tests such as sit-ups, push-ups, pull-ups, squat thrusts, and various strength tests. Arvey and Faley (1988) maintain that these tests are less likely to be legally supported because they do not represent “samples” of actual work behavior. This was especially evident in the classic 1982 New York City firefighter case, *Berkman v. City of New York*. The physical agility test items were selected using the constructs defined by Fleishman (Fleishman, 1964). None of the women tested passed the New York City firefighter test while 46 percent of the men did. The Court stated, “Nothing in the concepts of dynamic strength, gross body equilibrium, stamina, and the like, has such a grounding in observable behavior of the way firefighters operate that one could say with confidence that a person who possesses a high degree of these abilities as opposed to others will perform well on the job” (Arvey & Faley, 1988, p. 279).

Listed below are the decisions of the nine sex discrimination cases reviewed by Hogan and Quigley (1986). The common denominator of these cases was that each used common fitness items and, in some instances, in combination with work simulation tests such as a dummy drag. The ruling of only one case (*Hardy v. Stumpf*) supported the defendant’s use of the test. Following are the cases and the legal decisions—

1. *Hall v. White* — The physical agility tests (i.e., squat thrusts, sit-ups, push-ups, squat jumps, and pull-ups) were not found to be job related.
2. *Officers for Justice v. Civil Service Commission* — The physical agility tests, which were primarily upper body strength tests, did not predict job performance.
3. *Hardy v. Stumpf* — The tests were found reasonable, were supported by job analysis, and were not in violation of Title VII.
4. *United States v. City of Buffalo* — The tests used a weighted sum for height and weight and agility score. This method gave an advantage to taller persons, and the defendants were enjoined from further use of the method.
5. *Blake v. City of Los Angeles* — Job relatedness was not established for tests that combined running with job-related tests (e.g., scale 6-foot wall and drag 140-pound dead weight), and the validation studies were flawed.
6. *United States v. Philadelphia* — The tests (0.5 mile shuttle run, obstacle course, jump reaction time, and grip strength) did not show job relatedness.
7. *United States v. New York* — The tests were work simulations (e.g., shotgun aiming, tire change, etc.), but the job analysis was inappropriate for content validation and a different scoring strategy could have reduced adverse impact.
8. *Harless v. Duck* — Tests (i.e., push-ups, sit-ups, and standing broad jump) were not proved valid or job related, and job analysis did not specify amount of strength exertion required.
9. *Berkman v. City of New York* — The validation strategy was inappropriate, and should have used construct or criterion-related validity. The physical tests were dummy carry, grip strength, long jump, flexed arm hang, agility test, ledge walk, and mile run.



## Lifting

Some employers have attempted to argue that gender is a bona fide occupational qualification (BFOQ) for lifting and materials handling tasks. The defense for BFOQ allows for intentional classification of applicants or employees in the narrow circumstances in which such a classification is judged as a reasonable business necessity. If gender is a valid BFOQ, an employer can lawfully refuse to hire a person on the basis of his or her gender (Rothstein et al., 1999). The EEOC Guidelines interpret this defense very narrowly. Some examples of BFOQ based on gender are rest-room attendant and acting parts for male and female roles (Hogan & Quigley, 1986; Rothstein et al., 1999).

In response to the deleterious working conditions encountered by women during the industrial revolution, many states passed laws that “protected” women from physical labor. The courts have ruled that gender is not a BFOQ for lifting and materials handling physical tasks. The finding of the Supreme Court case, *Dothard v. Rawlinson* [435 U.S. 702, 98 S.Ct. 1370, 55 L.Ed.2d 657 (1978)], was that an employer cannot reject a female applicant who is capable of performing the job requirements solely because many other members of her gender group cannot do so (Hogan & Quigley, 1986; Rothstein et al., 1999). In the 1969 case, *Weeks v. Southern Bell Telephone & Telegraph Company*, the company did not consider a female employee’s bid for a job vacancy as switchman because the job had a 30-pound lifting requirement. Although Southern Bell contended that the switchman’s job was “strenuous,” the Court ruled that Southern Bell did not show that the job was so strenuous that most women could not perform it and that the lifting requirement was based on a “stereotyped characterization.” Southern Bell lost the case because they did not establish the validity of the lifting requirement (Arvey & Faley, 1988; Hogan & Quigley, 1986). The EEOC ruled (EEOC Decision No 71-1868 April, 22, 1971) that an employer’s use of only a few women as a small sampling of all women to perform lifting work was insufficient to establish that women were not qualified for jobs.

The courts have ruled that it is unlawful to disqualify women from being assigned to a job if an alternative exists with respect to the heavy lifting. In the case, *McLean v. State of Alaska* (Alas 1978, SCt. 18 EPD § 8787, 583 P2d 867), the task involved carrying 100-pound bundles of laundry. The Court ruled that an alternate was available by making up laundry in bundles less than 100 pounds. The position of the Court was that designating this job as a male job was unlawful. This position is consistent with the sound ergonomic practice of engineering excessive, physically demanding demands out of the job by redesigning the job (Waters, Putz-Anderson, Garg, & Fine, 1993). A problem, however, is that it often is not easy or even possible to redesign the job.

## Physiological Parameters

Strength, aerobic capacity, and body composition are physiological parameters that have been used for making employment decisions. This section briefly summarizes relevant cases. In addition to cases that were resolved by the courts, two additional EEOC cases that did not result in a final legal opinion are discussed.

**Submaximal  $\text{VO}_2\text{max}$  Test**—In a 1992 case involving a preemployment test for entry-level mill workers in a company's logging and sawmill operations,' physiological test principles helped decide the case. The preemployment test used by the timber company was ruled illegal because the test disproportionately excluded women qualified for the jobs. The test had three items: board pull ergometer to measure strength (pulling 30-, 50-, and 70-pound weights for specified durations), a 6-minute step test using an 11-inch bench; and a visual assessment of the applicant's gross body coordination. The legal problem was with the step test.

The step test required the applicant to wear a heart rate monitor. The test required that the subject exercise at an intensity of 10 metabolic equivalents (METs). The cut-score for the test was the physiological capacity to exercise < 85 percent of the applicant's heart rate estimated maximum aerobic capacity using a 10-MET power output. Since the female passing rate on the test was 42.4 percent of the male passing rate, the test showed disparate impact under the Uniform Guidelines 4/5s rule. The judge stated,

*Simpson has met its burden of showing that the test is job related and serves, in a significant way, the company's legitimate employment goal of hiring a physically fit workforce, in that those who pass the test, as a group, are likelier to be able to do the jobs adequately and safely than are those, as a group, who do not pass the test. The test nonetheless unnecessarily excludes qualified applicants, a disproportionate number of whom are women.*

The judge ruled that the methods used to administer a step test introduced gender bias, and the step test cut-score of 10 METs should be reduced to 8.5 METs, the cut-score used in testing existing employees who sought a transfer from one division of the company to another. The judge further ruled that part of the adverse impact was due to the failure to use an adjustable height bench or different height benches for men and women, and that this would not change the selection device except to increase accuracy. Lastly, part of the adverse impact was attributed to poor test administration. The judge stated that in some instances, women applicants were not given timely instructions about what they should eat or drink before the test; men and women were required to wait together while the test was administered; and female applicants were treated in ways that caused tension and anxiety, which could affect the outcome of the step test.

The use of heart rate-scored employment tests should be viewed with caution. A major problem with heart rate-monitored tests is that many applicants take medically prescribed medications that affect heart rate. For example, beta-blocker medication is commonly prescribed for hypertension. Although the medication is effective in lowering blood pressure, it also lowers resting, exercise, and maximum heart rate. The person's drug-affected maximum heart rate would need to be known to accurately determine exercise heart rate at a given percentage of maximum aerobic capacity. This would not likely be known.

**Aerobic Capacity**—The recent court ruling of *Lanning v. SEPTA* (U.S. 3<sup>rd</sup> Circuit 1999) is likely to have a major impact on the use of physiological variables to establish employment cut-scores. SEPTA is regional mass transit authority that operates principally in Philadelphia, Pennsylvania. A job analysis showed that SEPTA police officers had to chase suspects, and this often involved running up stairs. Subject matter experts (SMEs) were interviewed by an exercise physiologist to

determine the level of physical exertion necessary to perform tasks. The SMEs reported that a reasonable level of physical exertion was to run 1 mile in full gear in 11.78 minutes. The exercise physiology expert judged this as too low (i.e., 11.78 minutes per mile) and recommended the 1.5-mile run test with a more demanding cut-score of 12 minutes (i.e., 8.00 minutes per mile). This cut-score represented a  $\text{VO}_2\text{max}$  of 42.5 ml/kg/min, which was the same level recommended for selecting firefighters (Davis, 1992). The test was administered to incumbents, and the pass rates were 6.7 percent for women and 55.6 percent for men. The pass rate of women was only 12.1 percent of the men's, substantially lower than the 80 percent required by the 4/5s rule of the EEOC Guidelines.

Expert witness evidence for the defendant demonstrated that aerobic fitness was related and an important ability required by the transit police officer's job. Evidence was presented to show that the aerobic capacity of more than 52 percent of the persons arrested was 48 ml/kg/min, and only 27 percent of those arrested had an aerobic capacity of less than 42 ml/kg/min. Additional evidence was presented that showed a statistically significant correlation between aerobic capacity and arrests, arrest rates, and service-related awards. Although these data indicated that aerobic capacity was an essential determinant of job performance, the court ruled that a discriminatory cutoff score of the capacity to run 1.5 miles in 12 minutes ( $\text{VO}_2\text{max} = 42.5\text{ml/kg/min}$ ) is impermissible unless it represented the minimum qualification necessary for successful performance of the job in question. The legal foundation for this ruling was the Supreme Court interpretation from *Griggs v. Duke Power Co.* on the business necessity doctrine. The Court's interpretation was that a discriminatory cutoff score must be validated to show that it measures the minimum qualifications necessary for successful performance of the job. The Court further ruled that this was consistent with EEOC Guidelines that the cut-score "be set so as to be reasonable and consistent with normal expectations of acceptable proficiency within the work force." The Court went on to further declare that this is the only way to be certain to eliminate the use of excessive cutoff scores that have a disparate impact on minorities.

A stated goal of SEPTA was to respond to a perceived need to upgrade the quality of the police force. The Court indicated that there were three options open to help SEPTA achieve its stated goal of increasing the aerobic capacity of its police officers and to be consistent with Title VII. The options listed in the court ruling were as follows—

1. Abandon the test as a hiring requirement but maintain an incentive program to encourage an increase in the officer's aerobic capacities
2. Validate a cutoff score for aerobic capacity that measures the minimum capacity necessary to successfully perform the job and maintain incentive programs to achieve even higher aerobic levels
3. Institute a nondiscriminatory test for excessive levels of aerobic capacity, such as a test that would exclude 80 percent of men as well as 80 percent of women through a separate aerobic capacity cutoff for the different sexes.

Defining the aerobic intensity of work tasks is well founded in the discipline of exercise physiology and consistent with the *Lanning v. SEPTA* court ruling. Energy cost tables for common work tasks and recreational activities are published in several sources (Astrand & Rodahl, 1970;

Durnin & Rahaman, 1967; McArdle et al., 1991; Passmore & Durnin, 1955). These estimates are expressed in kilocalories per minute, oxygen consumption, or metabolic equivalents (METs).

A current, important research focus is to define the energy cost needed to fight fires. This research effort can be attributed to the amount of litigation leveled at the validity of firefighter pre-employment tests and the use of age to terminate employment. Several investigators (Barnard & Duncan, 1975; Davis & Dotson, 1978; Lemon & Hermiston, 1977; Manning & Griggs, 1983; O'Connell, Thomas, Caddy, & Karwasky, 1986; Sothmann, Saupe, Jasenor, & Blaney, 1992) published data showing that fire suppression work tasks have a substantial aerobic component. In an important study, Sothmann and a team of researchers (Sothmann et al., 1990) provide strong evidence that the minimum  $\text{VO}_2\text{max}$  required to meet the demands of fire fighting is 33.5 ml/kg/min. The authors used a work sample test involving seven job-related firefighter tasks. The sensitivity (percentage of correctly classified unsuccessful performers) and specificity (percentage of correctly classified successful performers) for a  $\text{VO}_2\text{max}$  cut-score of 33.5 ml/kg/min was 67 percent and 83 percent, respectively. Lowering the cut-score to 30.5 ml/kg/min, dropped the sensitivity to 25 percent and increased the specificity to 95 percent.

The  $\text{VO}_2\text{max}$  cut-score of 33.5 ml/kg/min/year was used to help decide *Smith v. City of Des Moines, Iowa* (U.S. District Court Southern District of Iowa-Central Division, 1995) firefighter case. The City of Des Moines requires that all firefighters be certified to wear a respirator (SCUBA). The city's pulmonologist, in cooperation with fire department personnel, developed a testing program on the appropriate level of cardiopulmonary fitness necessary for SCUBA certification. The standard consisted of two parts: First, the standard was met if the firefighter's FEV1 was  $\geq 70$  percent. If the FEV1 was below 70 percent, the firefighter would be required to take a maximum exercise test with a minimum cut-score of 33.5 ml/kg/min.

The firefighter alleged employment discrimination against the city of Des Moines because of his age. The Court ruled that the duties of a firefighter are inherently dangerous and that the fire department requires its firefighters to have the level of fitness needed to respond immediately and effectively to emergencies. Further, the Court held that the standards were reasonable and based on the demands of firefighting. The plaintiff claimed that the  $\text{VO}_2\text{max}$  standard violates the age discrimination prohibition because the likelihood of failure increases with age. The Court ruled that the 33.5 cut-score for an person of the plaintiff's age to be in "average" or "good" condition and was therefore not unreasonable.

Although the 33.5 ml/kg/min value reflects the level of aerobic fitness required to meet the physiological demands of firefighting, the cut-score often recommended is in the 40s (Davis & Dotson, 1992; Sothmann et al., 1990). The rationale used is that aerobic fitness declines with age. The cross-sectional rate of decline in  $\text{VO}_2\text{max}$  is about 0.4 to 0.5 ml/kg/min/year (Buskirk & Hodgson, 1987; Jackson et al., 1995; Jackson et al., 1996). Although the average decline is well documented, there is evidence that people vary considerably in the rate that their aerobic capacity declines with age. Both cross-sectional and longitudinal data (Jackson et al., 1995; Jackson et al., 1996; Kasch, Boyer, VanCamp, Verity, & Wallace, 1990) suggest that about 50 percent of the rate that aerobic capacity declines with age is due to differences in lifestyle. The rate of decline for those who remain aerobically active and maintain their level of body composition is estimated to be about 0.25 ml/kg/min/year compared with the average of about 0.5 ml/kg/min/year.

The logic of using a  $\text{VO}_2\text{max}$  cut-score in the 40s with young applicants is that firefighters would not have the physiological capacity to meet the firefighter demands as they age (Sothmann et al., 1990). The Lanning *v.* SEPTA and Smith *a.* City of Des Moines court rulings suggest that setting cut-scores above the 33.5 ml/kg/min may be legally hazardous. Setting firefighter aerobic capacity cut-scores higher than 33.5 ml/kg/min in an effort to account for potential aging effects increases the disparate impact for both sex and age discrimination. Table 7.2 illustrates the potential influence of aging and gender on cut-scores of 33.5 and 42.5 ml/kg/min. Using the data from two large NASA/Johnson Space Center samples (Jackson et al., 1995; Jackson et al., 1996), the average  $\text{VO}_2\text{max}$  for men and women for selected ages was estimated. Using the standard errors for regression models used to estimate the cross-sectional decline in  $\text{VO}_2\text{max}$  of men and women, and the normal curve, the proportion of men and women for selected ages who would pass cut-scores of 33.5 and 42.5 ml/kg/min was estimated. The male and female pass rates for each age were used to estimate the likelihood of disparate impact for gender. Table 7.2 shows that the pass rate for both men and women is inversely related with age. Obviously, the pass rate of both men and women is lower for each age group and cut-score, showing that both age and cut-score affect pass rates. More important, both age and cut-score affect the likelihood of adverse impact against women. Table 7.2 documents that as age and cut-score increase, the likelihood of a disparate impact for gender increases. On the basis of the Lanning *a.* SEPTA ruling and the trends shown in Table 7.2, it appears that setting aerobic capacity firefighter standards is an uncertain legal task unless a minimum job-related standard can be validly defined.

**Table 7.2 Average  $\text{VO}_2\text{max}$  (ml/kg/min) for males and females of selected ages, percentages that would pass the cut-score, and 4/5s rate for adverse impact**

	Average $\text{VO}_2\text{max}$		Cut-score 33.5 ml/kg/min		Cut-score 42.5 ml/kg/min	
	Female	Male	Females %Pass	Males %Pass	Females %Pass	Males %Pass
25	39.3	48.0	81.6	97.9	30.8	77.9
35	33.9	43.4	52.4	91.8	9.0	54.0
45	28.5	38.8	21.8	77.0	1.5	30.1
55	23.1	34.2	5.3	54.0	>0.1	12.1
65	17.7	48.0	0.7	29.1	>0.1	3.4

**Strength**—Hogan (1991) reports that the physiological construct related to most industrial tasks is strength. Not only is a sufficient level of strength necessary to perform many common industrial tasks, the lack of sufficient strength demanded by the task is likely to increase the **risk** of injury (Dehlin, Hendenrud, & Horal, 1976; Herrin, Jaraiedi, & Anderson, 1986; Magora, 1970; Snook, Campanelli, & Hart, 1978).

In the case EEOC *a.* Shell Western E and P (U.S. District Court, Central District of California) a preemployment strength test was challenged for sex discrimination. The test was used to screen applicants for entry-level jobs at oil and gas production facilities. The task analysis identified the physically demanding tasks to be push-and-pull forces required when completing well-pulling tasks. The amount of push-and-pull force required to perform the tasks was measured and

used to define the cut-score. The sum of grip, arm, and torso isometric strength was found to be highly correlated ( $\geq 0.80$ ) with work-sample push/pull tests. Regression models were used to define the level isometric strength needed to generate the level of push-and-pull force required by the task. This isometric strength level was used to define the cut-score (Laughery & Jackson, 1982).

The isometric strength tests were shown to have a disparate impact based on sex. This was expected and consistent with data showing that the strength of the average woman is about 50 percent of the strength of men (Baumgartner & Jackson, 1999; McArdle et al., 1991; NIOSH, 1977; Wilmore & Costill, 1994). A major concern raised by the plaintiffs' expert witness was that the isometric strength tests just measured upper body strength, and he argued that women performed work tasks like those identified by the task analysis with their legs. A major argument raised was the failure to also measure leg strength. In addition to failure to measure leg strength, evidence introduced by the plaintiffs suggested that the physical demands of the job changed. This resulted in an agreement between the EEOC and the defendant to redo the validation study and to include a leg strength test.

The task analysis of the new validation study documented that the physical demands of the job had changed (Jackson, Osburn, Laughery, & Sekula, 1998). The physically demanding well-pulling work was now being done by contract labor. The physically demanding tasks required by the current workers were valve cracking and lifting valves that weighed 75 pounds. The validation study was completed under the supervision of an exercise physiologist appointed by the EEOC. The study documented that isometric arm, shoulder, torso, and leg strength were correlated with valve cracking and lifting work-sample tests. The isometric strength tests were also found to be correlated with supervisor ratings of the worker's physiological ability to do the work. Simple linear and logistic regression models were used to define the minimum level of strength required to do these tasks. These data were used to define the cut-score. Post hoc analysis showed that all female incumbents exceeded the cut-score. Presently, the EEOC-supervised validation study has not been challenged.

**Body Composition**—A major issue in the case of the EEOC *v.* Mountain States Telephone and Telegraph Co., d/b/a U.S. West Communications (U.S. District Court for the District of New Mexico) was the use of skinfold fat to select workers for outdoor telephone craft jobs. An important physically demanding task of these craft jobs was pole climbing. The issues leading to the development of this study were the large differences between male and female workers in turnover and accident rates. After six months, 43 percent of the women left the outdoor craft jobs compared with only 8 percent of the men. An extensive job analysis showed that pole climbing was an essential, physically demanding work task. Accident data showed that women sustained substantially more injuries than men from falls while climbing or working on poles.

Using the results of a pilot study completed by exercise physiologists (Bernauer & Bonanno, 1975), industrial/organizational psychologists (Reilly, Zedeck, & Tenopir, 1979) completed a criterion-related validation study designed to develop a test for selecting applicants with the physiological capacity to climb poles safely. The four criteria of job performance were—

1. time to complete the pole-climbing school,
2. completion of pole-climbing school (a number withdrew from the course),
3. field observations of pole-climbing proficiency, and
4. accidents for six months after entering outdoor craft work.

A series of physical ability tests and the pole-climbing criteria were obtained on a sample consisting of 78 female and 132 male pole-climbing school applicants. Multiple regression selected a three-item battery consisting of body density estimated from skinfold fat, balance, and an isometric arm strength test. The statistically significant correlations between the three tests and the four criteria were time to complete the course, 0.46; training drop-out, 0.38; field observations of pole-climbing proficiency, 0.53; and accidents, 0.15.

The three-item test was used to select and disqualify applicants for pole-climbing school. Successful completion of pole-climbing school was a requirement for the outside craft position. In the validation study (Reilly et al., 1979), body density of the male and female subjects was estimated with gender-specific skinfold equations (Sloan, 1967; Sloan, Burt, & Blyth, 1962). A perceived limitation of these equations was the use of different combinations of skinfolds. In an effort to standardize test procedures, the final battery used just triceps skinfold because this site was common to both male and female equations. It is well documented in the literature that the women's triceps skinfold is significantly larger than that of men's. To illustrate, the means ( $\pm$  SD) for triceps skinfold of the women and men used to develop generalized skinfold equations (Jackson & Pollock, 1978; Jackson, Pollock, & Ward, 1980) were  $18.2 \pm 5.9$  for women; and for men,  $14.2 \pm 6.1$ . Using a common cut-score for men and women resulted in a disparate impact against women.

The argument made for the female plaintiffs was that the use of triceps skinfolds was discriminatory because of the gender-specific difference in triceps skinfold. Although it is physiologically sound to assume that persons with high levels of body fatness will have more difficulty climbing poles, and the statistical evidence of the criterion-related validation study supported this assumption, the defendants in the case agreed to stop using the test. Although this case did not progress to the point of a court ruling, it does illustrate the dangers of using body composition data for defining employment cut-scores. This agreement resulted in the development of new tests for the telephone industry that, at the time this chapter was written, has not been challenged in court.

---

## A View to the Future

With all the complexities that affect the U.S. judicial system, trying to determine what will happen in the future is very risky. Some recent U.S. Supreme Court decisions do suggest that discrimination as defined by ADEA and ADA is changing —

- \* The two recent Supreme Court cases of *Sutton v. United Airlines, Inc.*, and *Murphy v. United Parcel Service, Inc.*, showed that the burden of documenting a disability clearly rests with the employee and that an employee who fails to control a controllable disability may lose protection under ADA (Rothstein et al., 1999). A question may be raised whether fitness, especially weight or body composition, will be viewed as “controllable disability.”
- A recent U.S. Supreme Court ruling struck down part of a Federal law that allowed state employees to sue employers for age discrimination. The case was brought on by senior college professors who were paid less than their younger colleagues. The Court ruled that older

workers have no real constitutional protection against discrimination by state employees. Although the ruling does not seem to have a direct application to physical testing, the text from the court ruling suggests that it may have relevance to public-safety positions. A quote of the court ruling from the *Houston Chronicle* states, “As a result, states can discriminate against older workers merely by showing that the discrimination is rationally related to a legitimate state interest, the court said. For example, a state can have a mandatory retirement age of 50 for police based on its desire to ensure that all of its officers are in top physical condition, the court added” (*Houston Chronicle*, January 12, 2000).

## Endnotes

1. Findings of Fact and Conclusions of Law as to Liability, *EEOC v. Simpson Timber Company*, U.S. District Court, Western District of Washington, 1992.

---

## References

1. Arvey, R. D., & Faley, R. H. (1988). *Fairness in Selecting Employees*. (2<sup>nd</sup> ed.). Reading, MA: Addison-Wesley Publishing Co.
2. Astrand, P., & Rodahl, K. (1970). *Textbook of work physiology*. New York: McGraw-Hill.
3. Barnard, R., & Duncan, H. W. (1975). Heart rate and ECG responses of firefighters. *Journal of Occupational Medicine*, 17, 247–250.
4. Baumgartner, T. A., & Jackson, A. S. (1999). *Measurement for evaluation in physical education and exercise science*. (6<sup>th</sup> ed.). Dubuque: Wm. C. Brown.
5. Bernauer, E. M., & Bonanno, J. (1975). Development of physical profiles for specific jobs. *Journal of Occupational Medicine*, 17, 22–33.
6. Buskirk, E. R., & Hodgson, J. L. (1987). Age and aerobic power: the rate of change in men and women. *Federation Proceedings*, 46, 1824–1829.
7. Davis, P., & Dotson, C. (1978). Heart rate responses to fire fighting activities. *Ambulatory Electrocardiology*, 1, 15–18.
8. Davis, P. O., Dotson, C. O., O'Connor, J. S., & Confessore, R. J. (1992). *The development of a job-related physical performance test for Saint Paul firefighters*. Saint Paul, MN: Saint Paul Fire Department.
9. Dehlin, O., Hendenrud, B., & Horal, J. (1976). Back symptoms in nursing aids in a geriatric hospital. *Scandinavian Journal of Rehabilitative Medicine*, 8, 47–53.
10. Division of Industrial-Organizational Psychology, & Association, A. P. (1987). *Principles for the Validation and Use of Personnel Selection Procedures*. Washington, DC: Division of Industrial-Organizational Psychology, American Psychological Association.
11. Dunnette, M. D., Hough, L. M., (Eds.) (1991). *Handbook of industrial and organizational psychology*. (2<sup>nd</sup> ed.). (Vol. 2). Palo Alto, CA: Consulting Psychologists Press, Inc.



12. Durnin, J. V. G. A., & Passmore, R. (1967). *Energy, work and leisure*. London: Heinemann Educational Books, LTD.
13. Durnin, J. V. G. A., & Rahaman, M. M. (1967). The assessment of the amount of fat in the human body from measurements of skinfold thickness. *Journal of Applied Physiology*, 21, 681–689.
14. EEOC. (1991). Equal employment opportunity for individuals with disabilities: Final rule. *Federal Register*, 56, (144), 29 CFR Parts 1602 and 1627.
15. Fleishman, E. A. (1964). *The structure and measurement of physical fitness*. Englewood Cliffs, NJ: Prentice-Hall.
16. Golding, L. A., Meyers, C. R., & Sinning, W. E. (1989). *The Y's way to physical fitness*. (3<sup>rd</sup> ed.). Chicago, IL: National Board of YMCA.
17. Herrin, G. D., Jaraiedi, M., Anderson, C. K. (1986). Prediction of overexertion injuries using biomechanical and psychophysical models. *American Industrial Hygiene Association Journal*, 47, 322–330.
18. Hodgdon, J. A. (1992). Body composition in the military services: standards and methods. In B. M. Marriott & J. Gruinstrup-Scott (Eds.), *Body composition and physical performance: Applications for the military services* (pp. 57–70). Washington, DC: National Academy Press.
19. Hoffinan, T., Stouffer, R., & Jackson, A. S. (1979). Sex differences in strength. *American Journal of Sports Medicine*, 7, 265–267.
20. Hogan, J., & Quigley, A. M. (1986). Physical standard for employment and courts. *American Psychologist*, 41, 1193–1217.
21. Hogan, J. C. (1991). Chapter 11 Physical Abilities. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2<sup>nd</sup> ed., Vol. 2, pp. 743–831). Palo Alto, CA: Consulting Psychologist Press, Inc.
22. Jackson, A. S., Osburn, H. G., Laughery, K. R., Sekula, B. K. (1998). *Revalidation of methods for pre-employment assessment of physical abilities at Shell Western Exploration and Production, Inc., and CalResources LLC*. Houston, TX: University of Houston.
23. Jackson, A. S., Sekula, B. K. (1999). The influence of strength and gender on defining psychophysical lift capacity. *Proceeding of the Human Factors and Ergonomics Society*, 43, 723–727.
24. Jackson, A. S., Beard, E. F., Wier, L. T., Ross, R. M., Stuteville, J. E., & Blair, S. N. (1995). Changes in aerobic power of men ages 25–70 years. *Medicine and Science in Sports and Exercise*, 27, 113–120.
25. Jackson, A. S., & Pollock, M. L. (1978). Generalized equations for predicting body density of men. *British Journal of Nutrition*, 40, 497–504.
26. Jackson, A. S., Pollock, M. L., & Ward, A. (1980). Generalized equations for predicting body density of women. *Medicine and Science in Sports and Exercise*, 12, 175–182.
27. Jackson, A. S., Wier, L. T., Ayers, G. W., Beard, E. F., Stuteville, J. E., & Blair, S. N. (1996). Changes in aerobic power of women, ages 20 to 64 years. *Medicine and Science in Sports and Exercise*, 28, 884–891.
28. Kasch, F. W., Boyer, J. L., VanCamp, S. P., Verity, L. S., & Wallace, J. (1990). The effect of physical activity and inactivity on aerobic power in older men (a longitudinal study). *The Physician and Sportsmedicine*, 18, 73–81.
29. Laubach, L. L. (1976). Comparative muscular strength on men and women: A review of the literature. *Aviation, Space, and Environmental Medicine*, 47, 534–542.
30. Laughery, K. R., & Jackson, A. S. (1982). *Preemployment physical test development for land production and offshore drilling environments*. Houston: Shell Oil Co.

31. Lemon, P., & Hermiston, R. T. (1977). The human energy cost of firefighting. *Journal of Occupational Medicine*, 19, 558–562.
32. Magora, A. (1970). Investigation of the relation between low back pain and occupation. *Industrial Medicine and Surgery*, 39, 504–510.
33. Manning, J., & Griggs, T. (1983). Heart rates in fire fighters using light and heavy breathing equipment: similar near-maximal exertion in response to multiple work load conditions. *Journal of Occupational Medicine*, 25, 215–218.
34. Marriott, B. M., Grumstrup-Scott, J. (Eds.). (1992). *Body Composition and Physical Performance: Application for the Military Services*. Washington, DC: National Academy Press.
35. McArdle, W. D., Katch, F. I., & Katch, V. L. (1991). *Exercise physiology: Energy, nutrition, and human performance*. (3<sup>rd</sup> ed.). Philadelphia, PA: Lea & Febiger.
36. NIOSH. (1977). *Preemployment strength testing*. Washington, DC: U.S. Department of Health and Human Services.
37. O'Connell, E., Thomas, P., Caddy, L., & Karwasky, R. (1986). Energy costs of simulated stair climbing as a job-related task in fire fighting. *Journal of Occupational Medicine*, 28, 282–284.
38. Passmore, R., & Durnin, J. V. G. A. (1955). Human energy expenditure. *Physiological Review*, 35, 801–840.
39. Reilly, R. R., Zedeck, S., & Tenopyr, M. L. (1979). Validity and fairness of physical ability tests for predicting craft jobs. *Journal of Applied Psychology*, 64, 267–274.
40. Rothstein, M. A., Craver, C. B., Schroeder, E. P., Shoben, E. W. (1999). *Employment law*. (2<sup>nd</sup> ed.). St. Paul, MN: West Group.
41. Sloan, A. W. (1967). Estimation of body fat in young men. *Journal of Applied Physiology*, 23, 311–315.
42. Sloan, A. W., Burt, J. J., & Blyth, C. S. (1962). Estimation of body fat in young women. *Journal of Applied Physiology*, 17, 967–970.
43. Snook, S. H., Campanelli, R. A., & Hart, J. W. (1978). A study of three preventive approaches to low back injury. *Journal of Occupational Medicine*, 20, 478–481.
44. Sothmann, M. S., Saupe, K., Jasenor, D., & Blaney, J. (1992). Heart rate response of firefighters to actual emergencies. *Journal of Occupational Medicine*, 34, 797–800.
45. Sothmann, M. S., Saupe, K. W., Jasenor, D., Blaney, J., Donahue-Fuhrman, S., & Woulfe, T. (1990). Advancing age and the cardiorespiratory stress of fire suppression: Determining a minimum standard for aerobic fitness. *Human Performance*, 3, 217–236.
46. Vogel, J. A., & Friedl, K. E. (1992). Army Data: Body composition and physical capacity. In B. M. Marriott & J. Grumstrup-Scott (Eds.), *Body composition and physical performance: Applications for the military services* (pp. 89–104). Washington, DC: National Academy Press.
47. Vogel, J. A., Patton, J. F., Mello, R. P., & Daniels, W. L. (1986). An analysis of aerobic capacity in a large United States population. *Journal of Applied Physiology*, 60, 494–500.
48. Waters, T. R., Putz-Anderson, V., Garg, A., & Fine, L. J. (1993). Revised NIOSH equation for the design and evaluation of manual lifting tasks. *Ergonomics*, 7, 749–766.
49. Wilmore, J. H., & Costill, D. L. (1994). *Physiology of sport and exercise*. Champaign, IL: Human Kinetics.





# Appendix A

## Physical Fitness and Specific Health Outcomes

---

### Overweight and Obesity

Obesity has been defined as “A chronic disease characterized by an excessively high amount of body fat in relation to lean body mass” (Hoeger & Hoeger, 1998). The ACSM has offered the general criterion of “...a percent body fat that increases disease risk” (1991), which might be further clarified as “...weighing 20 percent more than recommended for a given body height” (Nieman, 1998). This begs the question of what defines the recommended weight in the first place. When using the widely accepted body mass index (BMI), calculated by dividing body weight in kilograms by height in meters squared ( $\text{kg}/\text{m}^2$ ), a measurement greater than or equal to 25 is classified as overweight and obese ( $\geq 30$ ) (NIH guidelines # 98–4803).

Unfortunately, the current lifestyle and environment of most Americans exacerbates the accumulation of excessive body fat. With multiple modern conveniences and transportation that requires little or no physical exertion, there are few daily chores that necessitate even moderate activity. Coupling inactivity with the high-fat, large-portioned convenience foods served at restaurants and **kiosks**, there is no mystery to the increasing body weight and BMI of our population. According to the U.S. Department of Health and Human Services, an estimated 107 million U.S. adults are overweight or obese and the numbers are increasing (2000). With more than half of the adults in the U.S. overweight or obese, it is important to consider the impact on overall health and morbidity and its implications for our society as well as for the quality of life of our citizens. Obesity has been linked to type II diabetes (West, 1978; Must et al., 1999; Hu, et al., 1999), cardiovascular disease (Lee et al., 1999; Hubert et al., 1983), hypertension (Bray, 1985), certain types of cancer and overall mortality (Lee et al., 1999). Obesity is considered a major contributor to many causes of disease and death, accounting for 15 percent to 20 percent of the annual mortality rate (Hoeger & Hoeger, 1998).

Using BMI to calculate obesity level, Must et al., (1999) described the relationship between the level of obesity and the prevalence of type II diabetes, gallbladder disease, coronary heart disease, high blood cholesterol, high blood pressure, and osteoarthritis. The study was a cross-sectional survey that used data from the Third National Health and Nutrition Examination Survey (NHANES III) that was conducted in two phases from 1988 to 1994. A total of 16,884 adults (25 years or older) were in the sample, all of whom were classified as obese (BMI greater than or equal to 25 kg/m<sup>2</sup>). From these survey demographics it was determined that 63 percent of men and 55 percent of women had a BMI of 25 or greater thus classifying them as overweight or obese. The results showed that “the prevalence of two or more health conditions increased with weight status category (overweight or obese) across all racial and ethnic subgroups” (Must et al., 1999). According to this study, the more overweight an individual was, the greater the chance of developing multiple health problems.

Together with its associated health problems, the 1995 price tag of obesity amounted to approximately \$99 billion in medical expenses and lost productivity (U.S. Department of Health and Human Services, 2000). That number is undoubtedly climbing. Such an extensive list of grave complications that arise with the incidence of obesity coupled with its medical costs and lost productivity demands serious attention be paid to identify and implement effective remedial measures.

Many studies have documented the salutary effects of regular and moderate physical exercise on obesity (DiPietro, 1995; Wilmore, 1996; Stefanick, 1993). According to *Physical Activity and Health: A Report of the Surgeon General*, “Several cross-sectional studies report lower weight, BMI, or skinfold measures among people with higher levels of self-reported physical activity or fitness” (U.S. Department of Health and Human Services, 1996). The Surgeon General also goes on to report the results of several comprehensive review articles and meta-analyses that examined the impact exercise training had on body weight and obesity.

These reviews conclude that—

1. physical activity generally affects body composition and weight favorably by promoting fat loss while preserving or increasing lean body mass;
2. the rate of weight loss is positively related, in a dose-response manner, to the frequency and duration of the physical activity session, as well as to the duration (e.g., months, years) of the physical activity program; and
3. although the rate of weight loss resulting from physical activity without caloric restriction is relatively slow, the combination of increased physical activity and dieting appears to be more effective for long-term weight regulation than is dieting done (p.134).

Researchers continue to delve into the causes and mitigation strategies for the prevalence and severity of obesity in our society. Genetic influences, high-calorie diets, and insufficient energy expenditure are the three central factors being explored. Genetic factors have been examined through studies on twins. Even when reared apart from one another, identical twins are much closer in body weight at middle age than are fraternal twins or siblings (Nieman, 1998). However, a study of 6,000 twins in Finland found that lifestyle factors were more important than genetics in determining weight gain over a six-year period (Korkeila, Kaprio, Rissanen, & Koskenvuo, 1995). The combination of these results indicates that the influence of lifestyle can be stronger than the influence of genetics in addressing body weight. It is widely known that a diet consisting of an

excessive number of calories and fat, particularly when combined with inactivity, contributes to weight gain. (Leibel, Rosenbaum & Hirsch, 1995).

Since it is known that the most effective intervention for obesity is the simultaneous increase of physical activity and reduction of caloric intake to a certain degree (Anderson et al., 1999), it may seem like there is a simple solution. However, there is more complexity to the formula than simply calories in and calories out. The complexity resides not only in determining what causes the body to expend the most calories but also in identifying the most effective approaches to encourage us to comply with these simple but effective interventions.

The basal metabolic rate (BMR) is the number of calories expended for basic bodily functions such as digestion, absorption, transportation of bodily fluids, as well as muscle and skin repair. It varies from individual to individual due to known factors such as genetics and body composition, and researchers continually seek other unknown determinants. So far, increasing muscle mass has been shown to increase the BMR and starvation can reduce it (Leibel et al., 1995). For the purposes of this chapter, the effect of physical activity on body-fat loss will be assessed for its effectiveness in reversing the problem of obesity.

An increase in muscle mass causes an increase in energy expenditure by an estimated 35 calories per pound of muscle per day (Campbell et al., 1994). This implies that more muscle results in greater caloric expenditure, which translates into less caloric storage in the form of body fat. Many studies have shown that regular physical activity maintains or increases lean body mass, expends energy, and helps to control weight (DiPietro, 1995; DiPietro, Kohl, Barlow, & Blair, 1999; Stefanick, 1993). In addition, an increase in muscle actually provides more fat-burning tissue and thus increases the baseline rate of caloric expenditure. The standard recommendation by government health experts and the ACSM is to engage in exercise at a moderate intensity for 30 to 60 minutes at least three days out of the week (Pollock et al., 1998).

Muscle or lean body mass is normally increased by strength training, which has become increasingly popular among health enthusiasts as scientists discover its many benefits including weight loss. Studies have shown that doing strength-training exercises, such as lifting weights, has a positive effect on bone density, preserving and increasing muscle mass, and reducing body fat. (Campbell, Crim, Young, & Evans, 1994). The current recommendation by the Surgeon General is to perform resistance-training activities at least twice per week. "At least 8 to 10 strength-developing exercises that use the major muscle groups of the legs, trunk, arms, and shoulders should be performed at each session, with one or two sets of 8 to 12 repetitions of each exercise" (U.S. Department of Health and Human Services, 1996). Similar guidance has been provided in the ACSM Fitness Prescription Position Stand (Pollock et al., 1998). It is clear that physical activity is central to obtaining positive results.

Knowing the results of the studies and their implications for better health and quality of life are merely the necessary conditions, not the sufficient conditions, for dealing with overweight and obesity. The chances that an obese person is willing or able to engage in the type, duration, and intensity of activity that is recommended are not good. A recent study surveyed almost 108,000 U.S. adults to examine the prevalence of attempts to lose weight and the strategies that were implemented (Serdula et al., 1999). The results showed weight loss as a common concern, with 28.8 percent of men and 43.6 percent of women trying to lose weight. However, only about one-fifth of the men and women reported actually implementing the recommendation to simultaneously decrease

caloric intake while engaging in regular physical activity (150 minutes per week). Even with such a large percentage of the population trying to decrease their weight, the mean body weight of U.S. adults has increased by 7.6 pounds during the past 15 years (Fine et al., 1999). The indications from these results are that people either do not know or do not adhere to the proven guidelines for weight loss. Individual compliance is the major factor here, as it is with all medical interventions, and it must be respected as a contributor to overall fitness.

The inability of Americans to maintain a healthy weight average results in diminished health-related quality of life and vitality, according to the investigation by Fine et al., (1999). Importantly, self-efficacy, image, and general well-being are associated with increased physical fitness. After surveying more than 40,000 adult women, investigators found that weight gain was associated with decreased physical function and vitality and increased bodily pain regardless of baseline weight. Weight loss on the other hand was associated with improved physical function. Even with the resulting improvements in health, most people in the United States are not taking action to eradicate the problem as evidenced by our increasingly overweight and obese population. It seems that a Catch-22 is in operation here. Because of the decrease in physical function caused by weight gain, it is difficult for many overweight people to be physically active enough to begin the weight reduction process. A high level of body fat makes activity more taxing, uncomfortable, and frustrating, which promotes their current sedentary lifestyle. It is an accelerating, downward spiral that results in further weight gain and increased susceptibility to illness and injury. The most important challenge to our nation's leading health experts is to first of all educate our people to the benefits of weight loss resulting from exercise in terms of longevity and quality of life.

Thus, it appears that promoting awareness of the efficacy of lifestyle activity as an alternative to structured aerobic exercise could be effective in controlling the rate of overweight and obesity mainly by making compliance with the intervention more palatable to the majority of the population at risk. Implementing the results of these studies that suggest lifestyle changes and small increments of activity throughout the day may encourage otherwise inactive people to begin to engage in a more physically oriented lifestyle. The problem of an overly fat nation is obviously significant, as it exacerbates multiple health conditions already present. If obesity is proactively addressed and eradicated, the onslaught of many health problems can be avoided. These are powerful findings in favor of adopting a set of activities that may easily become part of our daily activity repertoire.

---

## Hype

Blood pressure is "...the product of cardiac output and peripheral vascular resistance" (American College of Sports Medicine, 1994). This refers to the pressure exerted as the heart pumps blood through the veins, and is measured in milliliters of mercury (mm Hg). It is expressed in two numbers, systolic blood pressure (higher number) and diastolic blood pressure (lower number). The systolic pressure is exerted by the blood being forced against the walls of the arteries during the contraction of the heart. Diastolic pressure occurs during the relaxation phase of the heart, when the blood is again pushed against the artery walls. Hypertension is defined as a blood pressure reading of 140/90 or greater, with 160/100 being classified as severe.



Based on estimates made in 1996 by the American Heart Association (AHA), nearly 50 million Americans are hypertensive. Each year, high blood pressure kills over 37,000 Americans and it contributes to over 700,000 deaths (NIH, 1998). The National Heart, Lung, and Blood Institute reports that when left untreated, high blood pressure can—

- \* Cause the heart to get larger, which may lead to heart failure
- Cause small blisters (aneurisms) to form in the brain's blood vessels, which may cause a stroke
- Cause blood vessels in the kidney to narrow, which may cause kidney failure
- Cause arteries throughout the body to harden faster, especially those in the heart, brain, and kidneys, which can cause a heart attack, stroke, or kidney failure.

Studies have found that high blood pressure also affects the brain. When people have high blood pressure during middle age, they are more likely to experience cognitive problems 25 years later. This means that one's ability for memory, problem-solving, concentration, and judgment during old age is impaired (Launer et al., 1995). According to the ACSM, "...individuals with chronically elevated blood pressure have an increased probability of stroke, coronary artery disease, and left ventricular hypertrophy" (1996). Fortunately, high blood pressure can be controlled. Unlike obesity, effective medications can treat hypertension if taken daily. Unfortunately, these medications are expensive, can have adverse side effects, and require daily administration. To help prevent and control blood pressure, the National Institutes of Health (NIH, 1998) recommends that all people change their lifestyle behaviors in the following ways—

- \* Lose weight if overweight
- Reduce sodium intake to less than 2,300 mg per day
- Maintain adequate dietary potassium intake (fruits and vegetables)
- Limit alcohol intake
- Exercise regularly

NIH urges those who are hypertensive to implement the above recommendations for three to six months before starting drug therapy. Such recommendations are indicative of the powerful influence that lifestyle changes, including regular physical activity, can have on disease prevention. So far it has become clear that hypertension is a serious though preventable and treatable health issue. As with obesity, exercising at a moderate intensity for at least 30 minutes each day on most days of the week seems to have a positive effect on reducing hypertension.

The degree to which exercise can help in preventing and reducing hypertension has been shown in several studies (Blair et al., 1984; Folsom, Kushi, & Hong, 1996; Kokkinos, et al., 1998). Rueckert, Slane, Lillis, & Hanson (1996) has shown that there is a 20 to 50 percent greater risk for developing hypertension in inactive people when compared with those who are active. She collected data on 18 patients with high blood pressure before, during, and after they exercised. She found that walking on the treadmill for 45 minutes decreased their blood pressure below resting levels for up to two hours after they finished walking. In other research, Kokkinos et al., (1998) spent sixteen weeks examining 46 African-American men who were severely hypertensive and the effects of either anti-hypertensive medication alone or the medication combined with moderate exercise (1995). The researchers found

a significant decrease in the diastolic pressure of the exercising group, from 88 to 83 mm Hg., and an increase in the diastolic pressure in the medication-only group, from 88 to 90 mm Hg. ( $P=0.002$ ). They continued to monitor the subjects for 36 weeks and found substantial reductions in the diastolic blood pressure of the exercising group even after reducing their medications.

Because the research demonstrates a return to the pre-exercise blood pressure level soon after a person stops exercising, the ACSM recommends that hypertensives engage in frequent physical activity and incorporate this exercise as a permanent adjustment to their lifestyles. This recommendation is coupled with one that emphasizes aerobic activity rather than weight training. If weight training is to be used with aerobic exercise, the ACSM suggests a modification to the weight-lifting guidelines to include 10 to 15 repetitions (rather than 8 to 12) during each weight training exercise. Over time with continued activity, the blood vessels relax, creating a long-term lowering of the blood pressure. By adhering to the guidelines for regular, moderate aerobic activity, the research thus far seems to show that hypertensives can positively change their blood pressure and risk of mortality.

---

## Coronary Heart Disease and Cardiovascular Disease

Cardiovascular disease (CVD) is the number one killer in the United States today. There were an estimated 954,407 deaths in 1996 resulting from coronary heart disease (CHD) and stroke combined (NIH, 1996). This total is 41.2 percent of **all** deaths in that year, and its magnitude must result in a continuing and diligent investigation of approaches leading to its prevention and control. The Surgeon General reports that "...reviews of epidemiological literature have concluded that physical activity is strongly and inversely related to CVD risk" (1996). In other words, the more physical activity one engages in, the lower one's risk of CVD. In addition, the correlation between inactivity and CVD has been repeatedly examined, with findings of a direct relationship so convincing that inactivity is now listed as a risk factor for developing CVD (NIH, 1996).

There have been many studies examining the relationship between the dose-response relationship between exercise and risk of CVD (Kannel & Sorlie, 1979; Paffenbarger et al., 1984; Kannel, Belanger, Dagostino, & Israel, 1986). One such study at the Cooper Institute of Aerobics Research in Dallas, Texas, investigated the relationship between cardiovascular fitness levels and CVD. Included in this study were 25,341 male Cooper Clinic patients who underwent a maximal graded stress test and then were tracked for long-term follow-up. There were 226 cardiovascular deaths during the follow-up years. After accounting for other CVD predictors (high blood pressure, smoking, and high blood cholesterol), the researchers found a significant, inverse correlation between fitness levels and CVD in subjects with no other predictors ( $P=0.001$ ). The authors estimate that 20 percent of the 226 CVD deaths were attributed to low fitness level (Farrell et al., 1998).

The evidence repeatedly suggests that regular physical activity protects against the development of CVD (Bouchard et al., 1994; Haskell et al., 1992). This inoculation effect is due to the primary effects of exercise on improving cardiovascular health and to the favorable effects of physical activity on other CVD risk factors, such as high blood pressure, blood lipid levels, insulin resistance, and obesity (U.S. Department of Health and Human Services, 1996). Risk factors are defined as "...personal habits or characteristics that medical research has shown to be associated with an increased **risk**

of heart disease” (Nieman, 1998). In 1992 the American Heart Association added physical inactivity to the list of “major risk factors that can be changed.” The list also includes cigarette smoking, high blood pressure, and high blood cholesterol. The number of Americans with each of these three risk factors represents 20 percent to 25 percent of the population. However, the number of Americans with inactivity as a risk factor is 60 percent (American Heart Association, 1996).

By controlling the risk factors for heart disease it is thought that up to 90 percent of the occurrences of this disease could be prevented. Combining that conclusion with the results of numerous research studies on the positive health effects of exercise, one can deduce that regular physical activity could have prevented approximately 859,000 deaths in 1996. The question then is which activities are best suited for increasing cardiovascular health and preventing the risk factors for and the incidence of CVD. Although identifying the optimal types of exercises that are effective for different classes of individuals is very important, the most critical question is why most Americans are not exercising in light of its proven benefits. The answer lies in the psychology of compliance.

According to the collective findings reported by the Surgeon General, “Activity that reduced CVD risk factors and confers many other health benefits does not require a structured or vigorous exercise program. The majority of benefits of physical activity can be gained by performing moderate-intensity activities” (1996). These findings indicate that moderate-intensity activities confer significant health benefits, but such activity must be performed frequently. Fletcher et al. (2000) reported that the training effect of frequently engaging in activities such as biking, jogging, swimming, brisk walking, hiking, climbing stairs, aerobic exercise, tennis, soccer, and basketball, to name a few, are especially beneficial. When these activities make the heart rate exceed 40 percent to 50 percent of its maximal capacity, they are most effective. The activities that are considered low to moderate in intensity, ranging from 40 percent to 60 percent of maximum capacity include housework, gardening, dancing, and leisure walking. When performed daily, the health benefits of these activities are long term and predict a lower risk of cardiovascular disease (1996).

In a more specific examination of how much one needs to walk to reap its protective benefits, Sesso, Paffenbarger, Ha, and Lee investigated 1,564 women (mean age 45.5 years), initially free of CVD. The data were collected from 1962 until 1993. The authors looked at the calories expended in various activities such as number of stairs climbed, blocks walked, and sports played. They then divided those data into approximate thirds (<500, 500–999, 1000 or > kcal/week) to develop a quantitative dependent measure of fitness. During those years, 181 cases of CVD were identified. The researchers adjusted for other coronary risk factors and body mass index (BMI), and then compared the three “kcal expended” in terms of CVD risk. The results showed a 33 percent decrease in CVD risk for those women who walked at least 10 blocks per day (approximately 6 miles per week). In addition, there was an inverse association between lower BMI (<23 kg/m<sup>2</sup>) and CVD (1999).

When considering the broad and highly beneficial health effects of partaking in an active lifestyle, it is astounding to note that only 15 percent of adults performed the recommended amount of physical activity in 1997, and 40 percent of adults did not engage in any leisure-time physical activity (U.S. Department of Health and Human Services, 2000). Knowing the difficulty that many Americans have in meeting the 30-minute standard, three or more times per week, for general health and fitness, experts have offered an alternative approach (Pate et al., 1995). Their hypothesis is that more people will find ways to be active once they know that short periods of exercise (10 minutes)

a few times each day can protect against heart disease just as well as longer periods once a day. This approach may be more realistic to implement and has been proven to be effective.

---

## Diabetes

Diabetes is characterized by the inability of the human body to regulate its balance of glucose and insulin. The condition requires a consistent and stringent lifestyle that dictates specific eating times, type of diet, physical activity, blood glucose monitoring, and insulin injections as individually necessary. The people who are at the highest risk of developing diabetes are the ones with a high body mass index, especially if they are inactive. A study (Helmrich et al., 1991) examined nearly 6,000 men for 14 years, measuring their leisure-time physical activity (expressed as calories expended per week). The men who were both obese and inactive were four times more likely to develop non-insulin dependent diabetes (NIDDM) than the lean and active men were. In addition, the authors found that for each 500-calorie-per-week increase in activity expenditure, there was a 6 percent reduction in the risk of NIDDM.

The relationship between a sedentary lifestyle and the incidence of diabetes has been observed in other countries as they adopt Westernized or technologically advanced lifestyles. Those countries experienced major increases in the prevalence of NIDDM (West, 1978). Such findings are further supported by studies comparing individuals who moved from their native countries to more technologically advanced societies with their ethnic counterparts who remained in their homeland. The incidence of diabetes was much greater in those who moved (Ravussin, Valencia, Esparza, Bennett, & Schulz, 1994).

Another major six-year study of almost 7,000 Japanese-American men in Hawaii found that the rate of exhibiting the symptoms of NIDDM was lowest in the most active men, even after adjusting for obesity, age, family history, and other factors that contribute to NIDDM (Burchfiel et al., 1994). Those who were the least active had a 53.9 percent incidence rate, while the most active men had a 21.7 percent rate. This finding leads to the necessity to identify and quantify a threshold of activity that may protect against diabetes.

A group of scientists at Harvard conducted research to determine the benefits of moderate-intensity activity, such as walking, as opposed to vigorous activities (Hu et al., 1999) with regard to mitigation of diabetes. They examined this relationship through a prospective cohort study that included detailed data from more than 70,000 women in 11 U.S. states who were free of diabetes, cardiovascular disease, and cancer. The researchers got updates in 1986, 1988, and 1992. During the eight years of follow-up, 1,419 incidences of type II diabetes were reported. The results from this statistical analysis (adjusting for covariates) found that a faster-than-usual walking pace was independently associated with decreased risk. The researchers also discovered that equivalent energy expenditure, whether through walking or more vigorous activity, resulted in comparable magnitudes of risk reduction.

More research needs to be done for the evidence presented thus far to be substantiated. However, the link between obesity and diabetes is quite clear and there is an evident link between physical activity and obesity. With all of the research presenting the positive impact of moderate-

intensity exercise on these conditions, it is apparent that the current recommendation for regular activity may also decrease the risk for diabetes as well.

---

## Osteoporosis

Osteoporosis is characterized by a loss of bone mass, deterioration of bone tissue, increasing bone fragility, and increased likelihood of fractures. According to the Surgeon General, osteoporosis affects mostly older persons and is more common among women than among men (1996). This is due to the fact that women have a lower peak bone mass than men do, they lose bone mass at an accelerated rate when estrogen levels decline (usually after menopause), and they have a longer life span than men. In both men and women however, the three general reasons for developing osteoporosis are as follows: a deficient level of peak bone mass at physical maturity, failure to maintain this bone mass during the third and fourth decade of life, and the decline in bone mass that occurs during the fourth or fifth decade of life. The Surgeon General also reports that, "Physical activity may positively affect all three of these factors" (1996).

For bones to maintain their structure, they must have force applied to them. According to Dr. David Nieman, author of *The Exercise Health Connection*, "Healthy individuals who undergo complete bed rest for 4 to 36 weeks can lose an average of 1 percent bone mineral content per week, while astronauts in a gravity-free environment can lose bone at a monthly rate as high as one to four percent depending on the type of bone" (1996). Kirchmer, Lewis, and O'Connor (1996) conducted an extensive inquiry into the effects of exercise on bone mass and concluded that young adults who are athletic have a higher bone density than those who are sedentary. However, exercise is not the only factor contributing to bone strength. Hormones, diet, medications, disease, family history, race, gender, and age are all related to bone density. Similar to the salutary effects of exercise for other diseases, exercise may help to prevent or offset bone mass reduction no matter at what age one begins.

After conducting research with young adults, experts determined that physical activity plays a significant role in developing and maintaining bone mass. There also seems to be a compelling relationship between an increase in muscular strength and an increase in bone density. When young female athletes were tested for bone density, researchers found the greatest density for those who engaged in jumping and short bursts of powerful movement, such as one exhibits when playing basketball or volleyball. Interestingly, swimmers, who exercised in a weightless environment, had a very similar bone density compared with those who were sedentary (Nieman, 1996). Some researchers have observed a link between a history of lifelong physical activity and greater bone mineral mass as one's age advances (Snow, 1996). This positive effect of physical activity results in fewer incidences of hip fracture in older individuals.

Bone mass and strength naturally decline with age (Cummings 1985). Researchers also discovered that by the age of 90, one-third of all women and one-sixth of all men have sustained a hip fracture. The impact of hip fractures is extensive, accounting for more deaths, permanent disability, and medical institutional care costs than all other osteoporotic fractures combined (U.S. Department of Health and Human Services, 1996). The risk of falling combined with the impact of the fall and the strength of the bone are all factors determining hip fracture risk. Researchers

suggest that exercise may have a twofold effect on such a risk—decreasing the incidence and severity of falls (more muscle tone and perhaps better balance) and increasing the quantity and quality of mineral in the bones (Smith & Tommerup, 1995). Regardless of gender, age, or status, exercise reduces bone loss and increases bone mass, much like the effect of exercise on muscle.

The studies conducted on postmenopausal women conclude that bone mineral density is correlated with muscle strength (Sinaki, McPhee, Hodsdon, Merrit, & Offord, 1998). Unfortunately, the positive response of bone tissue to exercise is reversible, which indicates the need for continual activity throughout a person's adult life. Such activity should include a moderate amount of weight-bearing aerobic exercise and resistance training. A study conducted at Tufts University examined 39 postmenopausal women who engaged in intensive weight training for 45 minutes two times a week (Nelson et al., 1994). Compared with control subjects who were sedentary, the exercisers significantly improved their muscle mass as well as their bone density. The combination of hormones and exercise seems to have a very positive effect on bone density in postmenopausal women. This is illustrated by a study performed in Australia with 120 postmenopausal women. The researchers examined the forearm bone (a bone not affected by aerobic activity) in response to exercise and estrogen. The women who did aerobics only (without weight training) did not show any effect on bone density. However, the women who combined exercise and estrogen had significant improvements in bone density (Nieman 1998). Such results stress the importance of specificity relating to the impact of exercise to bone and the importance of hormones in overall bone development.

On the basis of many studies such as the ones described, the ACSM has deduced the following five principles of an exercise program to effectively prevent or treat osteoporosis—

- \* *Principle of specificity. If the leg bones are stressed by running and jumping, then the arm bones will not benefit unless they too are stressed with specific exercises (e.g., weight lifting).*
- *Principle of overload. For a bone to improve its density and strength, the exercise stress must exceed normal levels.*
- *Principle of reversibility. The positive effect of an exercise program on the skeleton will be lost if the program is stopped.*
- *Principle of initial values. People with the lowest levels of bone density and strength will experience more improvement from an exercise program than those with normal or above-normal bone density.*
- *Principle of diminishing returns. Each person has an individual genetic ceiling that limits the gains in bone mass. As the ceiling is approached, gains in bone mass will slow and plateau. (p. 3)*

If people participate in the recommended amount of activity, there can be a marked reduction in the prevalence and severity of osteoporosis.

---

## Cancer

Researchers have been investigating the benefits of exercise for preventing cancer for decades. There have been many studies that support an inverse relationship between exercise and some types

of cancer. The most positive results with exercise have been seen in the prevention of colon cancer and breast cancer. In fact, studies done by the Institute for Aerobic Research have shown that over an eight-year period, the overall cancer death rate was four times greater for physically unfit men than it was for the most fit men (Nieman, 1998). All types of cancer are thought to be 80 percent preventable through lifestyle and environmental factors, including diet, cessation from smoking and tobacco use, reduction of environmental hazards, and refraining from excessive alcohol use (Hoeger & Hoeger, 1998). Physical activity has many benefits but one interesting adjunct is the major change in the way people live when they are more physically active. By adopting a lifestyle that includes regular physical activity, people may be more likely to choose other healthy habits. These additional fitness-related behaviors, such as selecting a more healthy diet, will in turn reduce other risk factors and resolve other health problems (e.g., obesity). The combined effects of these positive habits may further reduce the likelihood of developing cancer.

The American Cancer Society and the National Cancer Institute have published a report stating that those living a healthy lifestyle have some of the lowest cancer mortality rates ever reported in scientific studies (American Cancer Society, 1986). There have been a number of additional studies correlating physical activity with some mitigating effects against cancer. A few of these studies have involved injecting mice with cancer-causing chemicals and then dividing the mice into exercise and non-exercise groups. The mice were then examined for the time and size of apparent cancer. In one particular study mice were consistently capable of clearing certain types of cancer after only nine weeks of exercise. Their inactive counterparts could not. This may be due to the ability of the body's macrophages to clear cancer cells for several hours after exercise.

The positive effects of exercise on the incidence and severity of breast cancer has been repeatedly studied. Sesso, Paffenbarger, and Lee (1998) studied 1,566 alumnae from the University of Pennsylvania who were cancer free between the years of 1962 to 1993. They established a physical activity baseline for each participant by asking the women which types of activities they engaged in and how often. They then divided the participants into three caloric expenditure categories; <500 kcal/week, 500–999 kcal/week, and >999 kcal/week. Their follow-up questionnaires found 109 cases of breast cancer during 35,365 person-years. After adjusting for age and body mass index (BMI), they discovered a significant effect of exercise on reducing the rate of breast cancer in postmenopausal women, but not in premenopausal women. They concluded that physical activity and breast cancer have a significant, inverse relationship among postmenopausal women.

Another study examined the responses of 25,624 women who filled out survey questionnaires about their leisure-time and work activity (Thune, Brenn, Lund, & Gaard, 1997). Over the course of 13.7 follow-up years, researchers identified 351 cases of breast cancer among the women. They found an inverse association between those engaging in more leisure-time activity and the incidence of breast cancer. In contradiction to the study cited above, Thune et al. found a greater reduction in risk among regularly exercising, premenopausal women than postmenopausal women, and in younger (<45 years of age) than in older women. In stratified analysis the risk of breast cancer was the lowest in lean women (BMI < 22.8) who exercised at least four hours per week. Those with a higher activity level also had reduced risk and the effect was again more pronounced among premenopausal women. The conclusion the researchers report is that physical activity, both during leisure time and at work, is associated with a reduced risk of breast cancer primarily in premenopausal women.

More than 30 studies have been published that have investigated leisure-time and work-time physical activity in relation to colon cancer. “Three fourths of these studies showed that physically active compared to inactive people have less colon cancer.” A frequent finding is that people who tend to sit for most of their workday or remain inactive in their leisure time have a 30 to 100 percent greater risk of contracting colon cancer (Nieman, 1998) than their more active counterparts. The Surgeon General reports on 18 studies conducted in a variety of populations, including China, Denmark, Japan, New Zealand, Sweden, Switzerland, Turkey, and the United States (1996). “Fourteen studies reported a statistically significant relationship between occupational physical activity and risk of colon cancer...” (p. 113). In eight study populations an inverse association was reported between physical activity and risk of colon cancer and results were usually consistent for men and women. Three studies that examined the effects of physical activity during early adulthood found no evidence to indicate that earlier activity did not affect risk of colon cancer later in life.

The number of different kinds of cancers and the difficulty in ascertaining a directly significant cause-and-effect relationship between activity and cancers makes it difficult to predict the protective effect of exercise. The studies on colon cancer and breast cancer clearly indicate that there is some link between consistent activity and a reduced incidence of cancer. The general guidelines to exercise moderately and regularly seem to provide at least some protection, especially when combined with other healthy habits such as a good diet and general health care.

---

## Clinical Depression

According to a report by the National Institute of Mental Health (1999), depression strikes more than 17 million Americans each year. This number is greater than the number of cases of coronary heart disease, cancer, or AIDS. The most troubling statistic is that 15 percent of depressed people commit suicide.

In 1996, the Surgeon General reported that “Epidemiological research among men and women suggests that physical activity may be associated with reduced symptoms of depression. In general, persons who are inactive are twice as likely to have symptoms of depression than are more active persons” (US. Department of Health and Human Services, 1996). Physical activity has been associated with improved mood and reduced anxiety right after and for up to several hours after an exercise session (Nieman, 1998). The implications for exercise in reducing depression include improved feelings of self-esteem, increased social interaction, relief from routine stresses, and brain chemical alterations. Any one of these factors or a combination of them may contribute to an enhanced mood state.

Approximately two-thirds of the people suffering from depression do not get professional help. There are a variety of possible reasons for such a lack of action. People may be too embarrassed or ashamed of feeling depressed, they may attribute their symptoms to other lifestyle factors such as poor diet, or they may feel too tired to bother. If left untreated, depression can result in years of misery and possibly self-inflicted injury or death. The cost of depression is quite high, and estimated \$43 billion per year due to lost work hours, lost productivity, and medical costs (Nordenberg, 1998). Although clinical depression may require more than one intervention (e.g., counseling and



antidepressant medication), exercise may be a simple and effective way to at least somewhat control depressive symptoms.

Although research has shown a connection between exercise and reduced feelings of tension and anxiety (APA, 1998), the question of how much exercise is effective for such results still remains. Because of the characteristic tiredness, lethargy, and disinterest in activity associated with depression, it is more probable that depressed persons would engage in minimal rather than vigorous activity for the relief of their symptoms. Many studies have examined the amount and type of exercise needed for decreasing depression. In a meta-analysis of 104 studies of 3,048 subjects, some very interesting dimensions of the effects of exercise on anxiety were documented and are summarized below —

- *Training programs usually need to exceed 10 weeks before significant changes in long-term anxiety occur.*
- *Exercise of at least 20-minutes duration seems necessary to achieve reductions in both present and long-term anxiety.*
- *Reductions in both present and long-term anxiety occur after aerobic but not anaerobic (e.g., weight lifting) exercise training programs. (p. 79)*

These findings indicate a need for prolonged and regular physical activity (American Psychological Association, 1998). Another research review suggested that “...exercise is an effective but underused treatment for mild to moderate depression” (Tkachuk, 1999). In this review, studies were analyzed from 1981 to the present. In each study, exercise was used as an intervention in treating psychiatric problems, including depression. The overall conclusion drawn from these studies was that non-aerobic forms of exercise, such as strength training, are as effective as aerobic exercise in treating depression. The studies also mention that “...less strenuous forms of regular exercise, such as walking, may be sufficient to demonstrate significant treatment effects” (Tkachuk, 1999). However, they note that more research is needed to confirm this finding.

There is presently sufficient evidence about exercise as a mitigating agent for depressive symptoms to support the government’s conclusion that regular physical activity enhances psychological well-being. Further, physical activity may even reduce the risk of developing depression, may reduce the symptoms of ongoing depression and anxiety, and may generally improve mood. Because of the complex factors influencing depression, a specific cause-and-effect relationship is difficult to establish. The research certainly indicates that physical activity may help the disorder and that a minimal amount of exercise will suffice.

---

## Stroke

Stroke is the common name for cerebrovascular disease or accident (CVA). A stroke occurs when arteries in the brain become narrow and eventually become clogged from atherosclerosis of the extracranial and/or intracranial arteries (U.S. Department of Health and Human Services, 1996). This buildup is similar to that which occurs in the heart to cause heart disease. High blood pressure is a major determinant of stroke occurrence. A stroke results in a deprivation of blood to the brain. This deprivation is extremely critical since the brain cells, which cannot store energy,

require 75 percent of the body's (resting) blood glucose to function. If the brain cells are deprived of blood for more than a few minutes, the cells die. The result is impaired vision, speech, motor function, comprehension, and possibly death (Nieman, 1998).

Each year approximately 500,000 Americans suffer from a stroke. The physical effects of a stroke are quite severe, resulting in death within one year for 30 percent of its victims, and within eight years for 60 percent (American Heart Association, 1997). Stroke ranks as the third leading cause of death following heart disease and cancer (NIH, 1996). Those who do not die from a stroke have a 50 percent chance of becoming functionally impaired enough to require assistance in caring for themselves. With such devastating effects, the prevention of stroke occurrence is a significant health and economic concern.

There are few studies that specifically relate physical activity to the primary prevention of stroke. However, the evidence is mounting to indicate that exercise is a significant primary preventative measure. Physical activity also has secondary beneficial effects on stroke by preventing or reducing the impact of other risk factors such as hypertension (Sacco et al., 1998). According to the American Heart Association (AHA), reducing the risk factors is the most effective way to prevent a stroke from occurring. AHA estimates that 70 percent of all strokes occur in people who have high blood pressure. The other major risk factors are cigarette smoking, excessive alcohol intake, and high blood cholesterol (AHA, 1996). Lifestyle factors are also linked to stroke indirectly. The finding that the rate of stroke occurrence fell 70 percent between 1950 and 1993 is attributed to changes toward a more active and healthy lifestyle that were occurring during this period. According to research into the effects of migration on health, men who were born in Japan (which has very high stroke death rate) and moved to California were found to have decreased their stroke death rate by 50 percent (Bronner et al., 1995).

The relationship between physical activity and many stroke risk factors (hypertension, obesity, diabetes, high blood cholesterol) is very strong. Thus, the current, general recommendations for exercise, emphasizing leisure-time physical activity, may be just as important in mitigating the risk of CVA as they are to other conditions. In order to examine directly the association between leisure time physical activity and stroke, Sacco et al., studied 369 subjects with a first stroke and 678 control subjects who were matched for age, sex, and race-ethnicity. The case subjects were interviewed within a median of 4 days after stroke onset. Each was asked to report the frequency and duration of 14 different recreational activities during the two weeks before the stroke. The researchers adjusted for cardiac disease, peripheral vascular disease, hypertension, diabetes, smoking, alcohol use, obesity, medical reasons for limited physical activity, education, and season of enrollment in the study. After these adjustments were made, a significant benefit of leisure-time physical activity was observed in all age, sex, and racial-ethnic groups. A positive dose-response relationship was found for both intensity and duration of physical activity as well. Simply stated this means that physical activity has a positive impact on reducing the risk of CVA no matter who you are, and that the more exercise you do, the better.

In addition, a study in Great Britain investigated the physical activity levels of 151 stroke patients and 161 controls (Shinton & Sagar, 1993). The researchers found a direct positive correlation between increases in duration of activity in the years before the study and an increase in protection from stroke risk. Risk of stroke dropped 56 percent in those who had engaged in regular and vigorous exercise between the ages of 15 and 25 with additional protection for those who con-

tinued exercising in adulthood. Shinton noted that vigorous exercise early in life seems to have a particular benefit, and that a lifelong exercise program offers the best health protection.

These studies support the beneficial effects of increasing activity as an effective countermeasure to stroke risk. Interestingly, other studies have found little or no additional benefit to high levels of exercise over moderate levels of physical activity, (Kiely et al., 1994). Although the evidence is still being analyzed in greater detail, the benefits of regular moderate exercise to reduce the risk of stroke is convincing. By reducing **risk** factors such as hypertension and obesity, which are clearly related to the incidence of stroke, a protective effect has been shown to occur to some degree. The preponderance of research findings seems to agree that some physical activity on a regular basis minimizes stroke risk. The current standards for health maintenance already incorporate recommendations for such moderate levels of physical activity and further research may provide even more details regarding the differential benefits of varying levels of exercise intensity.

---

## Musculoskeletal Problems

There are a multitude of problems and injuries that can occur because of weak muscles surrounding the joints. The musculoskeletal system has two kinds of connective tissue that support the joints: tendons, which link muscles to bones, and ligaments, which link bones to bones. The Surgeon General states that—

*Extensive animal studies indicate that ligaments and tendons become stronger with prolonged and high-intensity exercise. The effect is the result of an increase in the strength of the insertion sites between ligaments and tendons. These structures also become weaker and smaller with several weeks of immobilization, which can have implications for musculoskeletal performance and risk of injury. (p. 69)*

One of the most common injuries is low back pain, which affects an estimated 75 million Americans each year. Such widespread suffering is unnecessary, as 80 percent of the time it can be prevented (Hoeger & Hoeger, 1998). Low back pain is commonly caused by—

1. physical inactivity,
2. poor postural habits and body mechanics, and
3. excessive body weight.

Essentially, weak abdominal and back muscles, poor flexibility (especially of the lower back and hamstrings), and an abundance of fat lead to back problems. In theory, if the muscles of the abdomen and gluteal regions are not strong enough to support the spine and the weight surrounding it, then an unnatural forward tilt occurs in the pelvis. This tilt causes a curvature in the lower back and puts pressure on the spine, leading to low back pain. Sometimes the problem can be eradicated by stretching and strengthening tight and weak muscles.

The evidence is mixed, however, when it comes to poor musculoskeletal fitness as a predictor of back pain. Some studies have found weaker muscles correlating to lower back pain (Lee, Boreskie,

Law, & Russell, 1995), while others found little evidence of such a correlation (Ready et al., 1993). A cycle of inactivity is associated with back pain, thus creating weaker muscle that cannot effectively support the spine and may cause more problems that are more difficult to eradicate through movement. Other researchers (Malmivaara & Aro, 1995) have found that normal, moderate activity is most effective in treating back pain. This particular study assigned back-pain patients to either bed rest, regular activity, or back exercises. The patients who resumed ordinary activity recovered faster than those who stayed in bed or those who exercised.

---

## References

1. American Heart Association. (1996). Fletcher, G. F., Balady, G., Blair, S., Blumenthal, J., Caspersen, C., Chaitman, B., Epstein, S., Froelicher, S., Froelicher, V., Pina, I., & Pollock, M. (2000). *Statement on exercise: Benefits and recommendations for physical activity programs for all Americans*. American Heart Association. [On-line] <http://www.americanheart.org>
2. American Psychological Association. (1996). *Exercise gives a lift to psychotherapy. Psychology in daily life, get the facts*. APA Help Center. [On-line] <http://helping.apa.org/daily/exercise.html>
3. Anderson, R., Wadden, T., Bartlett, S., Zemel, B., Verde, T., & Franckowiak, S. (1999). Effects of lifestyle activity vs. structured aerobic exercise in obese women. *Journal of the American Medical Association*, 281, 335–340.
4. Blair, S. N., Goodyear, N. N., Gibbons, L. W., & Cooper, K. H. (1984). Physical fitness and incidence of hypertension in health normotensive men and women. *Journal of the American Medical Association*, 252, pp. 487–490.
5. Bouchard, C., Shephard, R. J., Stephens, T. (Eds). (1994). *Physical activity, fitness, and health: International proceedings and consensus statement, proceedings of the second international conference on physical activity, fitness and health*. Champaign, IL: Human Kinetics Publishers.
6. Bray, G. A. (1985). Complications of obesity. *Annals of Internal Medicine*, 103(6), pp. 1052–1062.
7. Bronner, Y. L. (1995). Session II wrap-up: Community-based approaches and channels for controlling hypertension in blacks: Barriers and opportunities. *Journal of the National Medical Association*, 87(6), Supplement, pp. 652–655.
8. Burchfiel C. M., Curb, J. D., Arakaki, R., Abbott, R. D., Sharp, D. S., Rodriguez, B. L., & Yano, K. (1996). Cardiovascular risk factors and hyperinsulinemia in elderly men: The Honolulu heart program. *Annals of Epidemiology*, 6(6), pp. 490–497
9. Campbell, W., Crim, M., Young, W., & Evens, W. (1994). Increased energy requirements and changes in body composition with resistance training in older adults. *American Journal of Clinical Nutrition*, 60, 167–75.
10. Cummings, S. R. (1985). Are patients with hip fractures more osteoporotic? Review of the evidence. *American Journal of Medicine*, 78(3), pp. 487–494.
11. DiPietro, L. (1995). Physical activity, body weight, and adiposity: An epidemiologic perspective. *Exercise and Sport Sciences Reviews*, 23, 275–303.

- 12 DiPietro, L., Kohl III, H. W., Barlow, C. E., & Blair, S. N. (1998). Physical fitness, physical activity, and functional limitation in adults aged 40 and older. *Medicine and Science in Sports and Exercise*, 30, p. 55–62.
13. Farrell, S. W., Kampert, J. B., Kohl, H. W., Barlow, C. E., Macera, C. A., Paffenbarger, R. S. Jr., Gibbons, L. W., & Blair, S. N. (1998). Influences of cardiorespiratory fitness levels and other predictors on cardiovascular disease mortality in men. *Medicine and Science in Sports and Exercise*, 9, 899–905.
14. Fine, J. T., Colditz, G. A., Coakley, E. H., Moseley, G., Manson, J. E., Willett, W. C., & Kawachi, I. (1999). A prospective study of weight change and health-related quality of life in women. *Journal of the American Medical Association*, 282 (22).
15. Fletcher, G. F., Balady, G., Blair, S., Blumenthal, J., Caspersen, C., Chaitman, B., Epstein, S., Froelicher, S., Froelicher, V., Pina, I., & Pollock, M. (2000). Statement on exercise: Benefits and recommendations for physical activity programs for all Americans. American Heart Association. [Online] <http://www.americanheart.org>
16. Folsom, A. R., Kushi, L. H., & Hong, C. P. (2000). Physical activity and incident diabetes mellitus in postmenopausal women. *American Journal Public Health*, 90, 134–138.
17. Helmrigh, S. P., Ragland, D. R., Leung, R. W., & Paffenbarger, R. S. (1991). Physical activity and reduced occurrence of non-insulin-dependent diabetes mellitus. *New England Journal of Medicine*, 325.
18. Hodgdon, J. A., Fitzgerald, P. I., & Vogel, J. A. (1990). Relationships between fat and appearance ratings of U.S. soldiers (Interim Report No. 90–1). San Diego, CA: Naval Health Research Center.
19. Hoeger, W. K., Hoeger, S. A. (1998). *Lifetime: Physical fitness and wellness*. Englewood, CO: Morton Publishing Company.
20. Hu, F., Sigal, R., Rich-Edwards, J. W., Colditz, G. A., Solomon, C. G., Willett, W. C., Speizer, F. E., & Manson, J. E. (1999). Walking compared with vigorous physical activity and risk of type 2 diabetes in women. *Journal of the American Medical Association*, 282.
21. Jones, B. H., Bovee, M. W., Knapik, J. (1992). Associations among body composition, physical fitness, and injury in men and women Army trainers. In B. M. Marriott, and J. Grumstrup-Scott (Eds.), *Body composition and physical performance, applications for the military services* (pp. 141–172). Committee on Military Nutrition Research, Food and Nutrition board, Institute of Medicine. Washington, DC: National Academy Press.
22. Kannel, W. B., Belanger, A., Dagostino, R., & Israel, I. (1986). Physical activity and physical demand on the job and risk of cardiovascular disease and death: The Framingham study. *American Heart Journal*, 112 (4), pp. 820–825.
23. Kannel, W. B., Sorlie, P. (1979). Some health benefits of physical activity. The Framingham study. *Archives of Internal Medicine*, 139 (8), pp. 857–861.
24. Kiely, D. K., Wolf, P. A., Cupples, L. A., Beiser, A. S., & Kannel, W. B. (1994). Physical activity and stroke risk: The Framingham study. *American Journal Epidemiology*, 140, 608–620.
25. Kirchmer, R. C., Lewis, R. D., & Oconor, P. J. (1996). Effect of past gymnastics participation on adult bone mass. *Journal of Applied Physiology*, 80, 255–332.
26. Kokkinos, P. F., Narayan, P., Collieran, J., Fletcher, R. D., Lakshman, R., & Papademetriou, V. (1998). Effects of moderate intensity exercise on serum lipids in African-American men with severe systemic hypertension. *American Journal of Cardiology*, 81 (6).

27. Korkeila, M., Kaprio, J., Rissane, A., Koskenvuo, M. (1995). Consistency and change of body mass index and weight. A study on 5967 adult Finnish twin pairs. *International Journal of Obesity Related Metabolism Disorder*, 19 (5), pp. 310–317.
28. Launer, L. J., Masaki, K., Petrovitch, H., Foley, D., Havlik, R. J. (1995). The association between midlife blood pressure levels and late-life cognitive function. The Honolulu-Asia aging study. *Journal of the American Medical Association*, 274 (23), pp. 1846–1851.
29. Lee, C. D., Jackson, A. S., & Blair, S. N. (1998). U.S. weight guidelines: Is it also important to consider cardiorespiratory fitness? *International Journal Obesity Related Metabolic Disorders*, 22 (Suppliment 2), S2–S7.
30. Leon, A. S., Connett, J., Jacobs, Jr., D. R., Rauramaa, R. (1987). Leisure-time physical activity levels and risk of coronary heart disease and death: The multiple risk factor intervention trial. *Journal of American Medical Association*, 258, pp. 2388–2395.
31. Leibel, R. L., Rosenbaum, M., & Hirsch, J. (1995). Changes in energy expenditure resulting from altered body weight. *New England Journal of Medicine*, 332, 621–628.
32. Ming, W., Kampert, J., Barlow, C., Nichaman, M., Gibbons, L., Paffenbarger, R., & Must, A., Spadano, J., Coakley, E., Field, A., Colditz, G., & Dietz, W. (1999). The disease burden associated with overweight and obesity. *Journal of the American Medical Association*, 282, 1523–29.
33. Must, A., Spadano, J., Coakley, E., Field, A., Colditz, G., & Dietz, W. (1999). The disease burden associated with overweight and obesity. *Journal of the American Medical Association*, 282, 1523–29.
34. National Institute of Mental Health. (1999). Depression Can Break Your Heart. Report No: 99–4592, [On-line] <http://www/nimh.nih.gov/publicat/heartbreak.cfm>
35. National Institutes of Health. (1998). *Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults: The evidence report*. NIH Publication No. 98–4083.
36. Nieman, D. C. (1998). *The exercise health connection*. Champaign, IL: Human Kinetics.
37. Nieman, D. C. (1999) *Exercise testing and prescription: A health-related approach*. 4<sup>th</sup> ed. Mountain View, CA: Mayfield Publishers.
38. National Institutes of Health Consensus Development Panel on Physical Activity and Cardiovascular Health. (1996). NIH Consensus Conference: Physical activity and cardiovascular health. *Journal of the American Medical Association*, 276.
39. Palmer, B., Rench, M. E., Carroll, J. W., & Constable, S. H. (2000). *Health and job-specific body composition standards for the US Air Force, Volume 1, Final Report* (CSERIAC-00-001A). Prepared for the Force Enhancement and Fitness Division, Department of Aerospace Physiology and Human Performance, USAF School of Aerospace Medicine.
40. Pate, R. R. (1988). The evolving definition of physical fitness. *Quest*, 40.
41. Pollock, M. L., Gaesser, G. A., Butcher, M. D., Despres, J., Dishman, R. K., Franklin, B. A., & Ewing, G., American College of Sports Medicine Positions Stand (1998). The recommended quantity and quality of exercise for developing and maintaining cardiorespiratory and muscular fitness, and flexibility in healthy adults. *Medicine and Science in Sports and Exercise*, 30 (6).
42. Ramsey, R. W. (1990). Fitness and warfighting: Combat troops should be physically fit—but can we be more specific? *Air Force Magazine*, 73 (4).
43. Ready, A. E., Boreskie, S. L., Law, S. A., & Russell, R. (1993). Fitness and lifestyle parameters fail to predict back injuries in nurses. *Canadian Journal of Applied Physiology*, 18 (1).

44. Recker, R. R., Davies, K. M., Hinders, S. M., Heaney, R. P., Stegman, M. R., & Kimmel, D. B. (1992). Bone gain in young adult women. *Journal of the American Medical Association*, 268.
45. Rueckert, P. A., Slane, P. R., Lillis, D. L., & Hanson, P. (1996). Hemodynamic patterns and duration of post-dynamic exercise hypotension in hypertensive humans. *Medicine and Science in Sports and Exercise*, 28 (1).
46. Serdula, M. K., Mokdad, A. H., Williamson, D. F., Galuska, D. A., Mendlein, J. M., & Heath, G. W. (1999). Prevalence of attempting weight loss and strategies for controlling weight. *Journal of the American Medical Association*, 282 (14).
47. Sesso, H. D., Paffenbarger, R. S., Ha, T., & Lee, I. M. (1999). Physical activity and cardiovascular disease risk in middle-aged and older women. *American Journal of Epidemiology*, 140 (4).
48. Shephard, R. J. (1992). A critical analysis of work-site fitness programs and their postulated economic benefits. *Medicine and Science in Sports and Exercise*, 24, pp. 354–370.
49. Sinaki, M., McPhee, M. C., Hodgson, S. F., Merrit, J. M., & Offord, K. P. (1986). Relationship between bone mineral density of spine and strength of back extensors in healthy postmenopausal women. *Mayo Clinic Proceedings*, 61.
50. Smith, E. L., Gilligan, C., Tommerup, L. J. (1994). Exercise and bone loss. *Advances in Nutritional Research*, 9.
51. Smith, R. (1995). Idiopathic juvenile osteoporosis: Experience of twenty-one patients. *British Journal of Rheumatology*, 34 (1).
52. Sterner, T. G. (1984). *Relationship between percent bodyfat as determined by visual estimation and underwater hydrostatic weighing*. Master's thesis, Springfield College, Springfield, OH.
53. Stofan, J. R., Dipietro, L., Davis, D., Kohl, H. W., & Blair, S. N. (1998). Physical activity patterns associated with cardiorespiratory fitness and reduced mortality: The Aerobics Center Longitudinal Study. *American Journal Public Health*, 88.
54. Tommerup, L. J., Raab, D. M., Crenshaw, T. D., & Smith, E. L. (1993). Does weight-bearing exercise affect non-weight-bearing bone? *Journal of Bone and Mineral Research*, 9 (8).
55. U.S. Department of Health and Human Services. (2000). *Healthypeople 2010*, 1 & 2. Washington, DC: U.S. Department of Health and Human Services.
56. Ward, G. M., Sutherland, T. M., Blanchard, J. M. (1976). *Evaluation of body composition of human subjects by means of visual appraisal* (Final Report Contract No. DAMD 17-74-C-4072). Colorado State University, Fort Collins, CO.
57. West, K. M. (1978). Diabetes in American Indians. *Advanced Metabolism Disorder*, 9.





# Index

## **AAHPERD test battery**

5

## **Absenteeism**

37, 50, 56, 141, 158

## **Activity**

5, 15, 18, 24, 27–47, 54–65, 70–74, 76, 78–83, 86–91, 94, 107–108, 119, 132, 136, 141, 203, 219, 223–241

## **Aerobic**

7–8, 12–13, 16–18, 23, 27, 29, 36, 38, 41, 44–48, 55–56, 58, 72, 78–83, 87, 101–110, 123, 125–132, 134–135, 137, 146–147, 162, 164, 172–174, 176, 179–181, 192, 199, 202, 211–215, 218–220, 226, 228–229, 232–233, 235, 238

## **Aerobic fitness**

12–13, 18, 36, 47–48, 82–83, 101–107, 109–110, 126–131, 137, 146–147, 162, 176, 199, 202, 213–214, 220

## **Aerobics**

12, 14, 25, 47–48, 65, 128, 228, 232, 241

## **Age**

4–6, 8, 10–13, 15, 17–19, 33, 35–36, 38, 43, 45, 48, 72–73, 80–81, 92, 104–110, 112–114, 117, 123, 129, 132–133, 135, 137, 175–176, 191, 195, 199, 201, 203–204, 206, 208, 214–215, 217–218, 220, 224, 227, 229–233, 236

## **Age discrimination**

201, 203, 208, 214–215, 217

## **Age Discrimination in Employment Act**

201, 203

## **Air Force**

1, 5–9, 11–14, 18–21, 23–25, 45, 47–51, 56, 58, 63–65, 94, 105, 115, 167, 172, 240

## **Air Force fitness**

11–13, 23, 58

## **Air Force Fitness Program**

11–13, 23, 58

## **American College of Sports Medicine**

5, 21, 29, 32, 39–40, 42, 50, 58–59, 63–64, 80, 94, 109, 116–117, 131, 180, 194–195, 197–198, 226, 240

## **American College of Sports Medicine Position Stand**

32, 40, 58

## **American Heart Association**

40, 43, 58, 60, 227, 229, 236, 238–239

## **Americans with Disabilities Act**

54, 180, 201, 203

**Appearance**

7, 14, 17, 19–20, 28, 49, 51–54, 56, 61, 73, 239

**Appearance in the military context**

51

**Army**

1, 5–11, 18–22, 24, 34, 50, 52, 59, 61, 63, 81, 91, 95–98, 115, 137, 167–168, 176, 193, 220, 239

**Army fitness**

9–10, 18, 21

**Army Fitness Program**

9–10

**Assessment**

1–5, 8–10, 14–15, 17, 20, 23, 25, 33, 47, 79, 88, 95, 112, 132–135, 173–174, 180, 185, 190, 193–194, 198, 212, 219

**Assessment procedures**

180, 185

**Basic ability**

101–102, 122–123, 125, 128, 131, 183, 193–194

**Basic ability tests**

101–102, 122, 125, 128, 131, 193–194

**Binet**

4

**Biomechanics**

88, 202

**Body**

5–9, 11, 13–17, 19–20, 22–25, 27, 29–30, 32–33, 35, 38–39, 41–42, 49, 51–54, 56, 59–65, 68, 71–72, 76, 79, 82, 84–88, 90–93, 96, 101–102, 106–107, 109–117, 120, 123–128, 130–137, 147, 150–151, 159–163, 166–168, 170, 173, 175–176, 194, 199, 205, 210–212, 214, 216–217, 219–220, 223–227, 229–230, 233, 236–241

**Body circumferences**

22, 114, 133

**Body composition**

6–7, 9, 11, 13, 15, 17, 19, 22–23, 25, 27, 29–30, 33, 38–39, 41, 49, 51–54, 56, 59, 61–63, 65, 96, 101–102, 107, 109–110, 112, 114–116, 123–128, 130–137, 147, 150–151, 161–163, 167–168, 170, 173, 175–176, 199, 211, 214, 216–217, 219–220, 224–225, 238–241

**Body composition equations**

115

**Body density**

110–114, 116, 134–137, 161, 173, 175, 217, 219

**Body fat**

5–8, 11, 13–15, 17, 19–20, 22, 24–25, 29, 42, 52–53, 64, 72, 107, 109–117, 120, 124–127, 133, 147, 150–151, 159–163, 170, 173, 205, 220, 223, 225–226, 241

**Body fat standards**

8, 13–15, 17, 19, 116–117

**Body fat, percent**

8, 11, 15, 17, 22, 25, 53, 64, 107, 109–117, 124–127, 133, 147, 150–151, 159–162, 170, 205, 223, 241

**Business necessity**

201, 204, 206, 209, 211, 213

**Canadian Standardized Test of Fitness**

5

**Cancer**

33, 38, 54, 63–65, 129, 134, 223, 230, 232–234, 236

**Capability**

68–69, 71–73, 89, 94, 132, 140–141, 173

**Cardiovascular disease**

33–34, 37–38, 42, 44, 54, 60–62, 64, 106, 115, 128–129, 132, 135, 155, 173, 223, 228–230, 239, 241

**Cattell**

3–4

**Centers for Disease Control**

8, 40, 42, 63, 65, 80

**Chemical plant workers**

165

**Circumferences, body**

22, 114, 133

**Civil Rights Act of 1964**

180, 201, 203

**Classification**

7, 24, 74, 80, 83, 86, 180, 189, 191, 211

**Clinical depression**

33, 55, 234

**Coast Guard**

8

**Common slope**

151

**Composition, body**

6–7, 9, 11, 13, 15, 17, 19, 22–23, 25, 27, 29–30, 33, 38–39, 41, 49, 51–54, 56, 59, 61–63, 65, 96, 101–102, 107, 109–110, 112, 114–116, 123–128, 130–137, 147, 150–151, 161–163, 167–168, 170, 173, 175–176, 199, 211, 214, 216–217, 219–220, 224–225, 238–241

**Content validity**

121–122, 139–140, 146, 181, 207

**Contingency tables**

179, 181–182, 188, 190, 192, 196–197

**Coronary heart disease**  
 33, 58, 62–63, 94, 104, 224, 228, 234, 240

**Court rulings**  
 208,215

**Criterion measures**  
 141,161,179,182,184

**Criterion-related validity**  
 139–141, 143–144, 146, 153, 166, 171, 179, 181–182, 186–187, 192, 207, 210

**Cut-point**  
 16, 27, 39, 48, 56

**Cut-score**  
 123, 141, 147, 150, 152–154, 158–159, 167, 180, 202, 207–208, 212–217

**Decision**  
 158–159,181,185,190,206,211

**Decision models**  
 185

**Demanding military jobs**  
 167

**Density, body**  
 110–114,116,134–137,161,173,175,217,219

**Department of Defense Directive 1308.1**  
 6, 24

**Department of Defense Instruction 1308.3**  
 6

**Diabetes**  
 30, 33–34, 54, 60–61, 64–65, 115, 128, 131, 223–224, 230–231, 236, 239, 241

**Discrimination litigation**  
 204

**Disparate impact**  
 201,204–209,212–213,215–217

**Diver training**  
 166–167,173

**Dose-response**  
 27, 32, 38–40, 45, 57, 61,224,228,236

**Electrical transmission lineworkers**  
 166

**Electromyography**  
 88–89, 93, 95

**Employee**  
 36–37, 55, 67–72, 74–78, 80–82, 84, 89–91, 130, 139, 146, 164, 176, 180, 197, 201–202, 204–205,208,211,217

**Employment**

54, 60, 91, 95, 120, 122–124, 139, 145–146, 149, 163, 169, 173–174, 179–181, 191, 193,  
197–198, 201–206, 208–209, 211–212, 214, 217, 219–220

**Employment law**

202, 205, 220

**Endurance**

3–4, 6–8, 13, 16, 18–19, 23, 29–30, 39–41, 46, 48–49, 56, 60, 63, 72, 101–102, 119–120,  
122–127, 130, 134, 137, 141, 143, 155, 161–163, 165, 167–168, 173, 175, 192

**Energy**

21, 32, 38–39, 44, 47, 59, 62, 71–72, 78–79, 81–83, 87, 93, 96–97, 105, 116, 123, 133,  
135–136, 147–148, 162, 173, 175–176, 192–193, 213–214, 219–220, 224–225, 230, 235, 238,  
242

**Environment**

30, 36, 51, 55–56, 67–68, 70–71, 85, 95, 122, 159, 184, 223, 231

**Equal Employment Opportunity Commission**

54, 60, 139, 180, 197, 201, 205

**Ergometer tests**

120

**Ergonomic data**

181, 185, 192, 197

**Error rate**

158

**Eurofit Test Battery**

5

**Evaluation**

1, 3, 5, 9, 15, 17, 19–20, 24, 60, 63, 65, 95, 97–98, 109, 132–135, 140, 142, 144, 146, 148,  
150, 152, 154, 156, 158–160, 162, 164, 166, 168, 170, 172–176, 178, 197–199, 218, 220,  
240–241

**Evaluation standards**

109

**Everyday work performance**

36

**Exercise**

4–5, 11–12, 15–16, 20–23, 27, 29–30, 32–48, 54–56, 58–65, 78–86, 94–98, 102–109, 117,  
128, 131–137, 143, 148–149, 162, 164, 166, 172–177, 197–199, 202, 207, 212–214, 216,  
218–220, 224–241

**Exercise prescription**

38–39, 45–46, 55–56

**Expectancy tables**

179, 181, 186–187, 196

**Fat, Body**

7

**Federal civil rights legislation**

203

**Field tests**

101–102,105,119–120,163

**Firefighters**

122, 127, 132–133, 135, 137, 161–163, 172–173, 175–176, 185, 192, 198–199, 202, 208, 210, 213–215,218,220

**Fitness**

1–25, 27–69, 72, 77–78, 80–83, 85, 89, 94–96, 101–107, 109–110, 114, 120–121, 124, 126–133, 135–137, 146–149, 151, 162–164, 167, 172–173, 176, 180, 191, 198–199, 202, 207, 209–210, 213–214, 217, 219–220, 223–226, 228–229, 237–241

**Fitness and readiness**

15

**Fitness standard**

13, 48–50, 56

**FITNESSGRAM**

5

**Fleishman**

3, 22, 75–76, 89, 91–93, 95–96, 143, 163, 173, 175, 193, 199, 210, 219

**Flexibility**

3, 6–7, 13–14, 23, 29–30, 39–41, 43, 63–64, 94, 101–102, 120–123, 130, 133, 163, 175, 191, 199,237,240

**Four-fifth's(4/5s) rule**

193,201,205,209,212–213,215

**General Health and Fitness Descriptors**

28

**General Sense of Well-being**

36

**Gilbreth**

2–3, 22

**Health**

4–7, 10, 14–16, 18–19, 21–25, 27–49, 51, 54–56, 58–65, 67, 69, 71–72, 76, 80, 85, 88, 90, 93–96, 101–102, 109–110, 112–113, 115, 118, 128–129, 131–133, 136–137, 139, 155, 175, 199,201,220,223–229,231,233–241

**Health and fitness**

28, 39–41, 44–45, 54, 56, 72, 229

**Health and safety**

67, 69, 88, 93

**Health benefits**

18,30, 33, 37, 39, 41–47, 51, 62, 229, 239

**Health-related**

6, 27–32, 34, 36, 38–42, 44–52, 54–56, 58, 60, 62–64, 66, 226, 239–240

**Heart disease**

33–35, 58, 62–63, 94, 104, 128–129, 131, 224, 228–230, 234–236, 240

**Heart rate**

4, 13, 29, 41, 72, 78, 80–82, 84, 91, 93, 97, 103–108, 132, 137, 149, 162, 164, 166, 172–173, 176, 212, 218, 220, 229

**Height**

4, 13, 17, 22–23, 29, 52, 72, 76, 78, 112–115, 120–121, 125, 133, 140–142, 156–157, 159, 161, 165, 167, 170, 192, 202, 205–206, 208–210, 212, 223

**Height and weight**

4, 23, 52, 112, 114, 202, 208–210

**Height and weight standards**

114, 202, 208–209

**Highway patrol officers**

163

**Hypertension**

33–34, 42, 54, 58–60, 62, 65, 128, 131, 204, 212, 223, 226–227, 236–239

**Individual differences**

1–3, 92, 120, 149, 169

**Injury rates**

130, 141, 158, 169

**Intercept differences**

150, 196

**Interrelated impact on conditions**

33

**Job**

6–7, 9, 13–14, 16, 19–20, 23–24, 27–29, 47, 51, 53–54, 56–57, 62, 67–82, 84, 86, 88, 90, 92–98, 100, 121–122, 125, 135, 139–148, 150, 152–154, 156, 158–170, 172, 174–176, 178–195, 197–198, 201–213, 216, 239

**Job analysis**

67–72, 74–78, 80, 82, 84, 86, 88, 90, 92–98, 100, 140–141, 145–146, 153, 160, 162–166, 181, 198, 210, 212, 216

**Job discrimination**

202–203

**Job performance**

7, 16, 23, 28, 53–54, 57, 67–68, 72–74, 77, 122, 140–141, 143, 145, 148, 154, 158, 161, 163, 166, 168, 179–182, 184–192, 194–195, 197, 207, 209–210, 213, 216

**Job performance criteria**

158, 168

**Job placement**

202

**Job specific**

19

**Job tasks**

9, 13, 76, 140–141, 146, 163, 179–182, 184–185, 192

**Job-specific fitness requirements**

50

**Laboratory**

14, 82, 101–104, 106, 108, 110, 112, 118–119, 123, 162–163

**Law**

23, 64, 73, 197, 201–203, 205, 207–208, 217–218, 220, 238, 240

**Legal**

57, 71, 146, 168, 179, 197–198, 201–206, 208, 210–216, 218, 220, 222, 224

**Legal issues**

198, 201

**Lifting**

14, 16, 29, 51, 57, 71–72, 76, 78, 90, 92, 94, 96–98, 108, 117, 119–120, 124–125, 130, 132–133, 135, 140–142, 152, 154, 156–157, 162, 164, 166–172, 174–176, 192–193, 197, 199, 205, 208, 211, 216, 220, 225, 232, 235

**Manual lifting tasks**

98, 154, 168, 176, 220

**Marine Corps**

6–9, 16–20, 24, 115

**Marine Fitness Program**

7–9

**Matching the worker to the job**

158

**Medical costs**

37, 224, 234

**Medical costs and health benefits**

37

**Metabolic fitness**

29–30, 41, 47

**Military body composition equations**

115

**Military fitness**

1–5, 17, 20, 27, 45, 51, 57

**Military readiness**

5, 45, 51, 56

**Military relevance**

49

**Morbidity**

29, 35, 44, 48–49, 54, 128, 223

**Mortality**

29, 33, 35, 37–38, 44, 48–49, 58, 60, 62–63, 65, 109–110, 128–129, 132, 135, 223, 228, 233, 239, 241



**Mortality and level of fitness**

37

**Muscle endurance**

18, 23, 63, 72, 101

**Muscle strength and endurance**

48

**Musculoskeletal**

29, 33–35, 55, 71, 76, 90, 96, 122, 128, 130, 159, 237

**Musculoskeletal problems**

35, 239

**Navy**

5–9, 13–25, 52–53, 81, 94, 114–115, 124, 133, 167

**Navy fitness**

14–15

**Navy Fitness Program**

14–15

**Normative data**

118, 179–182, 191, 196

**Norm-based**

1, 10

**Norms**

4, 13, 17, 90, 109, 118, 173

**Obesity**

14, 17, 19, 23, 32–34, 54, 56, 59–64, 113, 115, 128–129, 132, 134, 159, 176, 223–228, 230, 232–234, 236–238, 240, 242

**Occupational assessment**

1–3

**Occupational demands measurement**

1–2

**Occupational fitness**

1–3, 6–7, 11, 13, 15, 17, 20, 67–68, 72, 89

**Occupational standards**

7, 17, 20

**Occupational tasks**

22, 77, 80, 101, 173, 198

**Osteoporosis**

33, 42, 54, 60, 64, 231–232, 241

**Overweight**

33, 49, 56, 58–59, 61–63, 113, 115, 128–129, 223–228, 230, 232, 234, 236, 238, 240, 242

**Oxygen uptake**

5, 46, 48, 78–80, 82–83, 93, 97, 102–104, 123, 126, 133, 135–136, 147, 162, 164, 205

**Passing scores**

179–181, 185–187, 189, 191–193, 196

**Perceived exertion**

76, 80, 91, 93–94, 132, 164, 172

**Percent body fat**

8, 11, 15, 17, 22, 25, 53, 64, 107, 109–117, 124–127, 133, 147, 150–151, 159–162, 170, 205, 223, 241

**Performance**

2–3, 6–7, 9, 11, 13, 15–23, 27–32, 34, 36, 39, 45, 49–51, 53–55, 57, 61–63, 67–69, 71–75, 77–78, 81, 83–84, 88–89, 95–98, 101–102, 104–106, 108, 110, 112, 114, 116, 118, 120, 122–128, 130–141, 143, 145–146, 148, 151–152, 154, 158, 160–168, 171, 173–176, 179–202, 206–207, 209–210, 213, 216, 218–220, 237, 239–240

**Performance standards**

3, 7, 9, 16, 21, 23, 62, 179–180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200

**Physical**

1–25, 27–47, 49–65, 67–70, 72–73, 75–78, 80–88, 90–92, 94–98, 101–102, 104, 106–108, 110, 112, 114, 116, 118–122, 124, 126, 128, 130–140, 142, 144, 146–152, 154, 156, 158–164, 166–178, 180, 183, 186–188, 191–199, 202–205, 207–211, 213, 216–220, 223–231, 233–241

**Physical activity**

15, 27–45, 47, 54–65, 80, 82, 86, 107–108, 119, 132, 136, 219, 224–231, 233–241

**Physical Demands Analysis**

67–70, 72–73, 75–76, 86–88, 90–92

**Physical fitness**

1, 3–6, 8–17, 20–25, 27–30, 34, 37, 45–47, 49–54, 56–63, 65, 67–68, 80, 94–96, 101–102, 128, 131–133, 136, 163, 172–173, 191, 198, 219, 223, 226, 238–240

**Physical fitness assessment**

4, 8, 15

**Physical fitness measurement**

4

**Physical performance tests**

51, 57, 97, 101–102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 160–161, 163–164, 193, 197–199

**Physical testing**

159, 191, 193, 195, 202, 205, 208, 218

**Physical tests**

25, 139, 146–147, 158, 167, 171, 191–196, 199, 205, 208–210

**Physiological validation**

139, 143, 146, 148, 152, 156, 158–159, 171, 202

**Physiological parameters**

208, 211

**Physiological tests**

146–147, 153, 179

**Physiology**

23, 62–64, 79, 84, 94–98, 123, 131, 133–137, 148, 172, 175, 177, 197, 199, 202–203, 213, 218–220, 239–240

**Placement**

97,116,128,160,180,202

**Posture**

71–72, 74, 77–78, 88–90, 93

**Preemployment testing**

118,202,208

**Production rate**

158

**Protected group**

148,159,201,204–205,208

**Psychological tests**

139,146–147,153,171

**Psychometric tests**

203

**Psychometric test theory**

203

**Readiness and fitness**

49

**Recommendations from the American Heart Association**

43

**Recommendations, activity level**

40–44

**Regression slopes**

148–152

**Run field tests**

105

**Scores**

3, 10, 13, 15, 17–19, 72, 89, 122, 143–145, 147–148, 153, 155, 165, 179–182, 185–189, 191–197,213

**Scores, passing**

179–181,185–187,189,191–193,196

**Selection**

16, 20, 28, 47, 54, 56, 68, 70, 73, 76, 93–94, 97, 128, 139–146, 148, 150,152, 154, 156, 158–160, 162, 164, 166, 168, 170–174, 176–178, 180–181, 185–187, 189, 191–192, 197–199, 201,204–208,212,218

**§election method**

201,204,206–207

**Standards**

1–24, 26–28, 30–32, 34, 36, 38–40, 42, 44, 46–54, 56–58, 60–64, 66–69, 73, 82–84, 89–90, 95–96, 109–110, 113–114, 116–117, 133, 139, 144–147, 151, 153, 159, 164, 168–169, 172–173, 175, 179–182, 184, 186, 188, 190–192, 194–200, 202, 205, 208–209, 214–215, 219, 237,240

**Standards, body fat**

8, 13–15, 17, 19, 116–117

**Standards, fitness**

13, 48–50, 56

**Standards, height and weight**

114, 202, 208–209

**Standards, performance**

3, 7, 9, 16, 21, 23, 62, 179–180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200

**Steel workers**

163

**Strain**

67–69, 71–72, 77–78, 80–82, 84, 87–88, 93, 119

**Strength**

1, 3–11, 13–14, 16–20, 22–23, 27, 29–30, 36, 38–43, 46, 48–49, 51, 55–57, 63–64, 68, 72, 89, 92, 94–98, 101–102, 116–120, 122–128, 130–132, 134–136, 141, 143, 148–149, 151–158, 160–172, 174–175, 179, 181, 192–194, 199, 205, 209–212, 215–217, 219–220, 225, 231–232, 235, 237, 241

**Stress**

50, 67–69, 71–72, 77, 80–81, 84, 86, 91, 93, 98, 137, 146, 148, 158, 169–170, 173, 176, 199, 220, 228, 232

**Stroke**

33–34, 55, 58, 61–62, 64, 104, 115, 122, 129, 131, 227–228, 235–237, 239

**Submaximal aerobic**

104

**Success in training**

158

**Supervisory ratings**

158

**Task**

2–3, 7, 16, 29, 50, 53, 62, 67, 69–79, 81–82, 88–96, 102, 121–122, 128, 130–131, 136, 139–143, 146–156, 158–166, 169–171, 175, 183–186, 193, 205, 211, 215–216, 220

**Task performance**

16, 29, 69, 71–72, 77–78, 88–89, 95, 128, 152, 160–161

**Tasks, manual lifting**

168

**Task-specific**

27, 51, 57, 76

**Taylor**

2–3, 20, 23–24, 93–94, 96–98, 103, 136, 176, 181, 186–187, 196, 199

**Taylor-Russell Tables**

179, 184, 186–187

**Temperature**

68, 72, 79, 82, 84–86, 93, 96–97, 112

**Ternan**

4

**Test**

1, 4–5, 7, 9–18, 20–24, 45, 47, 49–51, 53, 56–57, 83, 89, 96–98, 101–108, 114–128, 130–133, 136, 139–154, 156–174, 176–182, 185–199, 201–210, 212–219, 228

**Test fairness**

148,179,193–197

**Test fairness**

148,179,193–197

**Test validation**

139,147–149,153,158–159,171,206

**Tests, physiological**

146–147,153,179

**Tests, psychological**

25

**Tests, psychometric**

203

**Thorndike**

3–4, 24

**Thresholds**

51, 96

**Title VII**

180,201,203–206,209–210,213

**Underground coal mining**

164

**Valid**

9, 18, 30, 52–53, 69–70, 87, 103, 105–106, 108, 121–122, 124, 126–127, 130, 140, 146–147, 152–154, 159, 161, 168–169, 180–181, 194–196, 201–202, 206, 208, 210–211

**Validation study**

145, 155, 159–160, 168, 185, 201–202, 207, 216–217

**Validation, physiological**

139,143,146,148,152,156,158–159,171,202

**Validity**

13, 22, 27, 57, 89, 92, 121–123, 126, 131, 134, 136, 139–141, 143–148, 153, 155, 162, 164, 166–167, 171, 174–176, 179, 181–182, 186–187, 192, 194–196, 199, 206–207, 210–211, 214, 220

**Validity, criterion-related**

139–140

**Work physiology**

79, 94, 123, 131, 172, 197, 203, 218

**Work sample tests**

101–102,121–125,130–131,161–164





2